

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:48:05 ; Search time 99.14 Seconds  
(without alignments)  
8086.296 Million cell updates/sec

Title: US-09-430-029-1  
Perfect score: 5828  
Sequence: 1 gatcatttcacaaatgcgc.....tcggcccggaacacgacgc 5828

Scoring table: IDENTIFY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/6.COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2490.6	42.7	9785	1	US-08-319-387-1
2	201	3.4	6379	1	Sequence 1, Appli
3	137.2	2.4	9785	1	Sequence 1, Appli
4	72.4	1.2	44377	3	US-08-319-387-1
5	72.4	1.2	44377	4	Sequence 7, Appli
6	72.2	1.2	3736	4	US-08-804-227C-7
7	69.4	1.2	20235	2	US-08-576-626A-1
8	65.4	1.1	1288	2	US-07-642-734C-3
9	60.6	1.0	1224	2	US-08-440-856A-9
10	60.6	1.0	1224	2	US-08-924-847A-1
11	60.6	1.0	1224	2	US-08-924-847A-3
12	60.6	1.0	1224	5	US-09-120-052-1
13	59.4	1.0	2634	5	US-09-120-052-3
14	59.2	1.0	30001	1	Sequence 1, Appli
15	59.2	1.0	30001	3	US-08-941-936-1
16	59	1.0	11219	2	US-08-125-468-1
17	58.8	1.0	1879	3	Sequence 1, Appli
18	58.8	1.0	1879	5	US-07-642-734C-1
19	58.4	1.0	1833	3	Sequence 1, Appli
20	58.4	1.0	1833	5	Sequence 5, Appli
21	58.2	1.0	1227	5	US-08-510-646B-5
22	58	1.0	4257	3	Sequence 6, Appli
23	57.4	1.0	2219	5	US-08-074-912-3
24	57.2	1.0	1227	2	Sequence 3, Appli
25	57.2	1.0	1227	2	Sequence 1, Appli
26	57.2	1.0	1227	5	US-08-510-646B-17
					Sequence 17, Appli
					Sequence 1, Appli
					Sequence 3, Appli
					Sequence 1, Appli

27 57.2 1.0 1227 5 US-09-120-249-3 Sequence 3, Appli

28 56.8 1.0 43280 3 US-08-804-227C-1 Sequence 1, Appli

29 55.8 1.0 13987 3 US-08-804-227C-13 Sequence 13, Appli

30 55.6 1.0 1957 1 US-08-295-060-3 Sequence 3, Appli

31 55.4 1.0 666 4 US-08-875-034A-1 Sequence 1, Appli

32 55.2 0.9 8438 1 US-07-945-283-1 Sequence 1, Appli

33 55 0.9 4776 4 US-08-852-401-1 Sequence 1, Appli

34 54.8 0.9 1146 2 US-08-482-385A-1 Sequence 1, Appli

35 54.8 0.9 2728 2 US-08-482-385A-5 Sequence 5, Appli

36 54.8 0.9 2943 1 US-08-042-747A-7 Sequence 7, Appli

37 54.4 0.9 9960 5 US-08-822-586-46 Sequence 46, Appli

38 54 0.9 1931 4 US-09-130-114-2 Sequence 2, Appli

39 53.6 0.9 30001 1 US-08-125-468-1 Sequence 1, Appli

40 53.6 0.9 30001 3 US-08-474-933-1 Sequence 1, Appli

41 53.4 0.9 1704 2 US-08-528-199-2 Sequence 2, Appli

42 53.4 0.9 1704 2 US-08-528-199-5 Sequence 5, Appli

43 53 0.9 3157 3 US-08-939-002A-1 Sequence 1, Appli

44 52.8 0.9 1700 6 PCT-US95-04801-1 Sequence 1, Appli

45 52.8 0.9 1771 3 US-08-533-669A-7 Sequence 7, Appli

## ALIGNMENTS

RESULT 1

US-08-319-387-1

; Sequence 1, Application US/08319387

; Patent No. 5543317

; GENERAL INFORMATION:

; APPLICANT: Shields, Malcolm S.

; APPLICANT: Francesconi, Stephen C.

; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,

; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Saliwanchik & Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/319,387

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/167,457

; FILING DATE: 15-DEC-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/694,718

; FILING DATE: 02-MAY-1991

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UWF-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 904-372-5800

; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9785 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-319-387-1





Db 7027 CCCTTGGCGAGGGTTCTGTCACGAGTCCGCAAAAGCTCATTTTCGAGAACAACTTTTCC 7086  
Qy 4716 gggaccagcgtaactgtgccccgccccgagatgatcgagcggtgcatcagcgctg 4775  
Db 7087 GGGCCAAAGCCATATTGTGCGGGCGCGCGGAATGATGACGCGCTGCTACGACCCCTG 7146  
Qy 4776 atgcaggggcgctgttcgagcgacatctatcacgagaagtctcatctcggcgccgac 4835  
Db 7147 ATCCAGGGCGGTTGTTCAGACCGGACATCTATTTTGAGAAAGTTCATTTCCGACCGCAT 7206  
Qy 4836 ggcgaacagacgcgagcccgctgttccggcggtgtga-catgagacgcggcgccgctat 4894  
Db 7207 GCCCAACAGTAGTCAGTCCGCTTTTCAAGAGGGTGTGATGCTACTCATCAGGACGGGA 7266  
Qy 4895 ggggagcggtcagatcgcgagacgagcgagcgatgctgctgctgctgctgctgctgctg 4954  
Db 7267 AGGTCTACGTGACCGCTCAAGCAGACCGCGGA-CGTTTTCAGCTGCGCTCTGGGTGAGTCTC 7325  
Qy 4955 tgcgtgccccgctggtggaactcggcgcgcgcttcctcgtgctgctgctgctgctgctgctg 5014  
Db 7326 TGCTTTCCGGATGGC--ACCTTGGCGTCCGGGATACCATGTTGTTGCTTAGCGGG 7383  
Qy 5015 ggtgccccgctgcaaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 5074  
Db 7384 GGTGGGAGTGTAAAGTTCGCGGTGTGACAGAGCGAGCGTGGCGCAAAAGTGGAGCAATGA 7443  
Qy 5075 ggcgtgccatgagcgcggaagagagacgagcgctgctgctgctgctgctgctgctgctgctg 5134  
Db 7444 GCGTGGCGACATTTCCGAGGCGGAAGAGAGCGCAAGGTGTGCTGCGGTGCCGGGTGG 7503  
Qy 5135 tgccgagcgcgctgcaactcgaactcgaactcgaactcgaactcgaactcgaactcgaactcga 5194  
Db 7504 CTCCTCAGTATGATGTGGNACTGGAGGTGTGCGCAAGATGCAAAAGCCCTTTTCAAG 7563  
Qy 5195 gcatggtgctgccccgagcgcgcgatcaacaagtaaccagagagagact-caccatgg 5253  
Db 7564 GGTTCAGCTTTTCAGGTAGCCCAAGATTCAACGAATTAAGCCGAAGAGATTAGACCATGA 7623  
Qy 5254 ggtgatcgctattggtcgtatgctgctgctgctgctgctgctgctgctgctgctgctgctg 5313  
Db 7624 GTGTGATCGAATCGCCATCGAGTCTGAAAGTGTGATGATGATGATGATGATGATGATGATG 7683  
Qy 5314 attacgtacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 5373  
Db 7684 ACTAGAAAACCTCT-GGAATGAAAAGGACGATGAGGACGAAACATGGAATGTGTACC 7742  
Qy 5374 tgaatgtggagcaatgggacaagtattcgtctgctgctgctgctgctgctgctgctgctgctg 5433  
Db 7743 TCAATGCTGGGACGAATGGGACAAAGTACTCGGTGATTCTGACGCGCTCCGATCAAGCGG 7802  
Qy 5434 gctcaagcatcgccctacaaggtcgagacgagcgagcgatgctgctgctgctgctgctgctg 5493  
Db 7803 GGTCTCAACCATGCTGCGCTCAAGGTTGAGCAGCATGCGGACCTGAATGCGTGCATAAGC 7862  
Qy 5494 gcatgaagcgtacggggtcgc 5536  
Db 7863 GPATGGAAGCTATGGATCAAGACACAAATGCTCCCCGAAGG 7905

## RESULT 2

US-08-499-215-1

; Sequence 1, Application US/08499215

; Patent No. 5612204

; GENERAL INFORMATION:

; APPLICANT: Saeki, Hisashi

; APPLICANT: Miura, Akira

; TITLE OF INVENTION: BIOLOGICAL DEGRADATIVE

; TITLE OF INVENTION: TREATMENT OF CHLORINE-SUBSTITUTED ETHYLENE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sugrue, Mion, Zinn, Macpeak &amp; Seas

; STREET: 2100 Pennsylvania Avenue

; CITY: N.W.

STATE: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
SOFTWARE: & WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,215  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP Hei-6-179689  
FILING DATE: 08-JUL-1994  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6379 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
STRAIN: No. 5612204ardia corallina B-276 (FERM P-4094; FERM  
BP-5124; ATCC 31338)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 910..1935  
OTHER INFORMATION: /product= "amoA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1935..2285  
OTHER INFORMATION: /product= "amoB"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2300..3802  
OTHER INFORMATION: /product= "amoc"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3805..4830  
OTHER INFORMATION: /product= "amod"  
US-08-499-215-1

Query Match 3.4%; Score 201; DB 1; Length 6379;

Best Local Similarity 54.9%; Pred. No. 7.4e-31;

Matches 546; Conservative 0; Mismatches 410; Indels 39; Gaps 6;

Qy 3857 cgaggaagacagacgagatgctgatgctgctgctgctgctgctgctgctgctgctgctgctg 3916  
Db 3849 CGAGGACGCGAGAGAGCCCTCTCGACGGCGCCCTGGCAACACAGCCCTGCTCTCTGAAGTACGG 3908  
Qy 3917 gtgctgtcacgggtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3976  
Db 3909 GTGCAAGCACGGGGCTCGGGGACCTGCAAGGTCTCGGCTGCTCGAGGGGACGTAGAGGA 3968  
Qy 3977 gggcgatgcgaacccgttcgctgctgctgctgctgctgctgctgctgctgctgctgctg 4036  
Db 3969 ACCCGGGTGG---TCGTTCCGCTGACGCGGAGGACCGGAGACGAGTGTATCTCTCC 4025  
Qy 4037 gtgctgtcgagcgtgcagcgacacacgtgatcgagcgcgcgcgcgcgcgcgcgcgcgcgc 4096  
Db 4026 GTGCGCCAGGTGCGCGCTGGAAACCGTGCACCTGACGCTGAGCGGAGCGGCGCTCACGGA 4085  
Qy 4087 tgcggaatcatcccggtc-----agggaacttcgccccgcgcgcgcgcgcgcgcgcgcgc 4147  
Db 4086 GGAGGAGTCTCTCTCGGGGACACCTCGCGGAGTCTCCAGACGGTCTGCGGGCGGTGCGA 4145  
Qy 4148 acagctcagcgaccacataaagtcgctgctgctgctgctgctgctgctgctgctgctgctg 4201  
Db 4146 GTTTCTACGGCGACATCGCCCGGTCTCGGCTCCGGCTAGAGCGGCGGCGGAGATCGC 4205





; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: ASCII(DOS) Text only  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/804,227C  
 ; FILING DATE: February 21, 1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Plant, Thomas, G.  
 ; REGISTRATION NUMBER: 35,784  
 ; REFERENCE/DOCKET NUMBER: X-8231  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 317-276-2459  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 44377 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 350..14002  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 14046..20036  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 20110..31284  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 31329..36071  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 36155..41830  
 ; US-08-804-227C-7

Query Match 1.2%; Score 72.4; DB 3; Length 44377;  
 Best Local Similarity 43.5%; Pred. NO. 1e-05;  
 Matches 328; Conservative 0; Mismatches 426; Indels 0; Gaps 0;  
 QY 987 gcccagagcccgagctgctgagggcgcccaaggcgacacctggaccgagccgagccgagcctggca 1046  
 |||||  
 Db 28626 GCCCGGGAGGCGCGGAGCGCGCCGCTGTGTGTGGTCCACCGCGCGCGGTGCGCGT 28685  
 QY 1047 gccctgccccgctacgtctgaggaacacgctgctgctgccccgacatccgctgagctgttcat 1106  
 |||||  
 Db 28686 ACGCCGACGACGCGGAGTGGAGCGCGCCGCGCGCGCGGTGGGGCGCTGCTGCGCTC 28745  
 QY 1107 gcgcgagaacctcgctcgacggcctgtgtatccgctctacgacgcttctgtcga 1166  
 |||||  
 Db 28746 CCGCGAGCGCGGAGCGCGCGCGCTGCTGTGTGACCTGGAGCGAGCGACCGCGCGCTC 28805  
 QY 1167 cgaacgagatcgctcgagggcggtcgagctgctgagccgagcttcatcccca 1226  
 |||||  
 Db 28806 CGCCCGGGCGGTGACCGAGCGCGCTCGCTCCGCGAAGCGAGCGCGGTCCGGGCGCG 28855  
 QY 1227 atggcacaccgagtgtaaccgctgtagcagcggtctgtaagacgattggccgcgaatc 1286  
 |||||  
 Db 28866 GACGGTGTACGTGCCCGCGCTGGAGCGCGCGCGCGCGCGCGCGCGCGCTCACGCC 28925  
 QY 1287 cgacgacaacgcggtgctgctgccccgtggacacgagctggttcgcgcgccgagac 1346  
 |||||  
 Db 28926 GCCCGACGACGCGTGTGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCTCGC 28985  
 QY 1347 ggcactggcagcgtggcgccgagcgctgacagatccgagggcgcgcgcgctcgacga 1406  
 |||||  
 Db 28986 CCGTGTGCTCGC 29045  
 QY 1407 agtgcgagcagttccacgacgagcgagggcgaggtcgagctcgctctgacgagcgga 1466  
 |||||  
 Db 29046 ACGCCCGCGCGCGGTCAACTTCCGCGACGCGCGCTCATCGCCCTCGGCATGTACCGGGCGA 29105

QY 1467 atctcccttaaccagaatgcccagatgtccaacgtattcatcgcttttcaggccaa 1526  
 |||||  
 Db 29106 GCGGAGATGGGAGACGGAGCGCGCGCGCACCGTCTCGAGGTGCGCCCGCGCGGTACCCGG 29165  
 QY 1527 tgagactccagacccgagatctgtgatgctgctgcccgaaccccgcgcggtgtgtgt 1586  
 |||||  
 Db 29166 TGTGCGCGTGGCGACCGCGTCTCGGCCCTGTGGGACGGCGCGCTGGGCGCGCTGTGGGT 29225  
 QY 1587 cgagtcgccccgcatggttaagatgacgagcgccgagacgctgacgagatccgcgcgaaac 1646  
 |||||  
 Db 29226 GCGCGACACCGCGCTGCTCGCCCGCTCCCGACGCGCTGCTACGCCACGCGCGCTC 29285  
 QY 1647 gatcgagaaactgacgagcagcgcttcgacctgacagctccaggtccaaactaatcac 1706  
 |||||  
 Db 29286 GGTCCCCCGGTGTCTCAGCGCGCTACTACGGTCTGGTACCCCTGGCGCGCTCAGGCC 29345  
 QY 1707 gctgtcagggccacatcgacgaggaagcagcagag 1740  
 |||||  
 Db 29346 GGGGAGCGGGTGTCTGTCACGCGCGCGCGGG 29379  
 RESULT 5  
 US-08-804-198-1  
 ; Sequence 1, Application US/08804198  
 ; Patent No. 5945320  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgett, Stanley G.  
 ; APPLICANT: Kuhstoss, Stuart A.  
 ; APPLICANT: Rao, Nagaraja R.  
 ; APPLICANT: Richardson, Mark A.  
 ; APPLICANT: Rostock, Paul R., Jr.  
 ; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PAUL R. CANTRELL 1138  
 ; STREET: LILLY CORPORATE CENTER  
 ; CITY: INDIANAPOLIS  
 ; STATE: IN  
 ; COUNTRY: USA  
 ; ZIP: 46285  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Macintosh  
 ; OPERATING SYSTEM: Macintosh 7.0  
 ; SOFTWARE: Microsoft Word 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/804,198  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CANTRELL, PAUL R.  
 ; REGISTRATION NUMBER: 36,470  
 ; REFERENCE/DOCKET NUMBER: P9113  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 317-276-3885  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 44377 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 350..14002  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 14046..20036  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 20110..31284  
 ; FEATURE:



















[illegible]

Search completed: September 26, 2000, 19:23:33  
Job time: 12928 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:47:32 ; Search time 5060.92 Seconds  
(without alignments)  
2055.201 Million cell updates/sec

Title: US-09-430-029-1  
Perfect score: 5828  
Sequence: 1 gatcatttcataatgcgc.....tcggccgggcaacacgacgc 5828

Scoring table: IDENTRY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba1:\*
  - 2: gb\_ba2:\*
  - 3: gb\_om:\*
  - 4: gb\_ov:\*
  - 5: gb\_pat:\*
  - 6: gb\_ph:\*
  - 7: gb\_pl1:\*
  - 8: gb\_pl2:\*
  - 9: gb\_pr1:\*
  - 10: gb\_pr2:\*
  - 11: gb\_pr3:\*
  - 12: gb\_ro:\*
  - 13: gb\_sts:\*
  - 14: gb\_sy:\*
  - 15: gb\_un:\*
  - 16: em\_fun:\*
  - 17: em\_hum1:\*
  - 18: em\_hum2:\*
  - 19: em\_in:\*
  - 20: em\_om:\*
  - 21: em\_or:\*
  - 22: em\_ov:\*
  - 23: em\_pat:\*
  - 24: em\_ph:\*
  - 25: em\_pl:\*
  - 26: em\_ro:\*
  - 27: em\_sts:\*
  - 28: em\_sy:\*
  - 29: em\_un:\*
  - 30: em\_vi:\*
  - 31: gb\_htg1:\*
  - 32: gb\_htg2:\*
  - 33: gb\_in1:\*
  - 34: gb\_in2:\*
  - 35: em\_ba1:\*
  - 36: em\_ba2:\*
  - 37: em\_hum3:\*
  - 38: em\_hum4:\*
  - 39: gb\_pr4:\*
  - 40: gb\_htg3:\*
  - 41: gb\_htg4:\*
  - 42: gb\_htg5:\*
  - 43: gb\_htg6:\*

- 44: gb\_htg7:\*
- 45: em\_htg1:\*
- 46: em\_htg2:\*
- 47: em\_htg3:\*
- 48: em\_hum5:\*
- 49: gb\_pl3:\*
- 50: gb\_pr5:\*
- 51: gb\_htg8:\*
- 52: gb\_htg9:\*
- 53: gb\_htg10:\*
- 54: gb\_htg11:\*
- 55: gb\_htg12:\*
- 56: gb\_htg13:\*
- 57: gb\_htg14:\*
- 58: gb\_in3:\*
- 59: gb\_htg15:\*
- 60: gb\_htg16:\*
- 61: gb\_htg17:\*
- 62: em\_htg4:\*
- 63: em\_htg5:\*
- 64: em\_htg6:\*
- 65: em\_htg7:\*
- 66: em\_hum6:\*
- 67: gb\_htg18:\*
- 68: gb\_htg19:\*
- 69: gb\_htg20:\*
- 70: gb\_htg21:\*
- 71: gb\_htg22:\*
- 72: gb\_htg23:\*
- 73: gb\_htg24:\*
- 74: gb\_htg25:\*
- 75: gb\_htg26:\*
- 76: gb\_htg27:\*
- 77: gb\_htg28:\*
- 78: gb\_htg29:\*
- 79: gb\_htg30:\*
- 80: gb\_htg31:\*
- 81: gb\_vil:\*
- 82: gb\_vil2:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %		Length	DB ID	Description
		No.	Match			
1	2738	47.0	6298	1	AB006479	Comamonas
2	2681.6	46.0	7823	1	AB024741	AB024741 Comamonas
3	2490.6	42.7	9785	5	124403	Sequence 1
4	2244.8	38.5	5029	1	PSETBMAF	L40033 Pseudomonas
5	1988	34.1	9511	2	AF026065	AF026065 Ralstonia
6	1949.2	33.4	8654	1	AB031996	AB031996 Ralstonia
7	1729.2	29.7	6112	1	PSEPHAEA	D28844 Pseudomonas
8	1708.8	29.3	6343	1	PPPEHYD	X80765 P. putida ge
9	1617.2	27.7	5449	1	PSEPHHYD	M60276 Pseudomonas
10	1616.6	27.7	5700	1	PPPHH	X79063 P. putida ge
c	1213.4	20.8	2934	2	AF012632	AF012632 Ralstonia
	1043.4	17.9	1117	1	AB016861	AB016861 Burkholder
12	799.6	13.7	5552	1	D85083	D85083 Acinetobact
13	779.2	13.4	5905	1	ACPHENOL	Z36909 A. calcoacet
14	740.8	12.7	1154	1	AB017631	AB017631 Comamonas
15	723.8	12.4	1127	1	AB016863	AB016863 Comamonas
16	707.4	12.1	1133	1	AB016860	AB016860 Pseudomon
17	665.6	11.4	1144	1	AB016859	AB016859 Pseudomon
18	646	11.1	1143	1	AB016857	AB016857 Pseudomon
19	643	11.0	1139	1	AB016854	AB016854 Pseudomon
20	617	10.6	985	1	AB016858	AB016858 Pseudomon
21	455.4	7.8	3511	2	AF112137	AF112137 Burkholder
22	451.8	7.8	2640	2	CSU93090	U93090 Comamonas s
23	431.2	7.4	3103	2	AF109307	AF109307 Pseudomon

25	405.4	7.0	2191	1	PSETBUD	M98806 Pseudomonas	CDS
26	393.2	6.7	614	1	AB016862	AB016862 Comamonas	
27	376.6	6.5	615	1	AB016856	AB016856 Pseudomon	
28	349.2	6.0	646	1	AB016855	AB016855 Pseudomon	
29	334.2	5.7	684	2	AF228346	AF228346 Acinetoba	
30	330.6	5.7	1145	1	PPTDNC	X59790 P.putida td	
31	265.6	4.6	1575	2	BP2020258	U02058 Pseudomonas	
32	261.2	4.5	1436	2	BCU47111	U47111 Burkholderi	
33	254.8	4.4	2433	2	AF112136	AF112136 Burkholde	
34	236.2	4.1	1146	2	U01826	U01826 Pseudomonas	
35	201	3.4	6379	1	RERAMOA	D37875 Nocardia co	
36	201	3.4	6379	5	E08269	E08269 gDNA of alk	
37	201	3.4	6379	5	I36963	I36963 Sequence 1	
38	201	3.4	6379	23	E10607	E10607 Alkenemooo	gene
39	201	3.4	6379	23	E10980	E10980 Gene of alk	
40	173.8	3.0	1575	2	BP2020258	U02058 Pseudomonas	CDS
41	152	2.6	8620	2	AF065891	AF065891 Ralstonia	
42	142	2.4	9037	1	PWWXYL	M64747 Pseudomonas	
43	141.8	2.4	209	1	AB011577	AB011577 Unidentif	
44	140.2	2.4	209	1	AB011575	AB011575 Unidentif	
45	138.6	2.4	209	1	AB011572	AB011572 Unidentif	

ALIGNMENTS

RESULT	1				
AB006479	AB006479	6298 bp	DNA	BCT	06-NOV-1998
LOCUS	Comamonas testosteroni gene for phenol hydroxylase component, ferredoxin like protein, catechol 2,3-dioxygenase, complete cds.				
DEFINITION					
ACCESSION	AB006479				
VERSION	AB006479.1	GI:3845544			
KEYWORDS	aphB; aphQ; aphP; aphO; aphN; aphM; aphL; aphK; catechol 2,3-dioxygenase; ferredoxin like protein; phenol hydroxylase component.				
SOURCE	Comamonas testosteroni (strain:TA441) DNA.				
ORGANISM	Comamonas testosteroni				
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.				
AUTHORS	1 (bases 1 to 6298)				
TITLE	Arai,H.				
JOURNAL	Direct Submission				
	Submitted (11-AUG-1997) to the DDBJ/EMBL/GenBank databases.				
	Hirotsuki Arai, The RIKEN Institute, Laboratory of Microbiology; Hirotsawa 2-1, Wako, Saitama 351-0198, Japan				
	(E-mail:harai@postman.riken.go.jp, Tel:048-467-9545, Fax:048-462-4672)				
REFERENCE	2 (sites)				
AUTHORS	Arai,H., Akahira,S., Ohishi,T., Maeda,M. and Kudo,T.				
TITLE	Adaptation of Comamonas testosteroni TA441 to utilize phenol: organization and regulation of the genes involved in phenol degradation				
JOURNAL	Microbiology 144 (Pt 10), 2895-2903 (1998)				
MEDLINE	99018839				
FEATURES	Location/Qualifiers				
source	1..6298				
	/organism="Comamonas testosteroni"				
	/strain="TA441"				
	/db_xref="taxon:285"				
	283..474				
gene	/gene="aphK"				
	283..474				
CDS	/gene="aphK"				
	/codon_start=1				
	/transl_table=11				
	/product="phenol hydroxylase component"				
	/protein_id="BAA34169.1"				
	/db_xref="GI:3845545"				
	/translation="MEYSTLDLPVCDLTRRSRVLRKRVGFVDFEFSVGWPELTVEL AMTEADFEAFCCQVQAKLL"				
gene	516..1508				
	/gene="aphL"				





[illegible]

D	b	3030	GCCTTTTATAGCAAGACCTTGCCCATGCTGTGCACCACCTGCCAGATCCTATGGCGTT	3089
Q	y	3032	cacgagcgggcgaaccgcagcaagaatcggcgcgcgcaaatcgaaactacctcgcgcaaca	3091
D	b	3090	CACCGAGCCGGTGACGCCACCAAGATCTGCTATCGCGAGTCGCACATTACGAGGCAACA	3149
Q	y	3092	gttcacttctgcagcgaccactgcaaggacatatcttgatacagagccgcgagaaaatcgt	3151
D	b	3150	GTACCACCTTCTGACGGCACGGCTGCAAGCATGCTTCCAGCATGAGCCCGAGAAGTTGT	3209
Q	y	3152	gcaggcgtggtctccggtgcaccagatccaatcaggcaactgcttcccgccgatcgga	3211
D	b	3210	CCAGTCTGGTTGCCCGTGACCAGATTATACAGGGCAACTGCTTAAGGCCGACGTGGA	3269
Q	y	3212	ccccggcgagggcttcogtatcgctcgccgcggtgctcgactactacgcggtgacgat	3271
D	b	3270	CCCCACGGCCGAAAGCTTTGACCCGCTGGCTGCGTGCTGGACTGGTACGGCATGAATGC	3329
Q	y	3272	ggggccgcgaacactcogatcttcgacggctcggaaacagaaacttcgcggcgtagcg	3331
D	b	3330	GGSCCGCGACAACATGGATTTTAGGGTTTCGGAGGATCAGCGCAATTTTCGGCGCTCGCG	3389
Q	y	3332	cgccgagggcacgcgcgaactgacccgcgaacagcaagaacattcttgacgagggccgcggaag	3391
D	b	3390	CGSGCATCCGCC-----ACCGGCAACTCGGCCAACAGCGACGCCAGTAGTAGCGCTTC	3440
Q	y	3392	cqccgatgcgaacacggggcgacaggagagaaaaatggcctcatcgctcgaacccc	3451
D	b	3441	CGCAAACTCAGCCAGAGGCC-----CAGCCATGAGGTGTCACCGCACATCAAGCCC	3491
Q	y	3452	tacgaactcccggtgaaggatgcgctcgagaagtcttcggcgccgctgctctacgtgtc	3511
D	b	3492	TATCGCTTTCCGCCCAAGGATGTGCGCGAGAACTTCCCGCGCCTCTGCTCTATATCGCG	3551
Q	y	3512	tgggaaaaacatctgatgtttcccgcgccgctctgctgcgctgcgcgccgacatgcgcg	3571
D	b	3552	TGGGAAGATCACCTGATGTTCTGCTCGCCGCTGGCTCTGCGCTGCGCGGCGGATATGCCT	3611
Q	y	3572	ttcgcgcgctggcggcgcacgtgctgcgcgcgcgctctacggtatcaccccgacttcgcg	3631
D	b	3612	TTGCGTGGCGTGAACCGGCTGTCGTGCCCGATGTTATGGCGCCACCCCGACTTGCA	3671
Q	y	3632	aagatcgaactgggatcgctcgagtggttcocggtcgggcgagcgtggcgccgagcccg	3691
D	b	3672	CGCATCGATGGACCGAGGTGAGTCAGTGGTTCAGTCTGGCGCCAGCCCTGGTGCGCCGATGCG	3731
Q	y	3692	gcgaagagcctggcgccgaacggcctcgggcaaggaactgatcagcttccgcacgccc	3751
D	b	3732	TCCAAGTCGTGSCCGACAAACGGCCTGACGCACAAGGACGTGATCCGCTTTCGCAACCCCG	3791
Q	y	3752	ggcctcgaacgcttcggcgggcgagcgttctgacgc-----	3788
D	b	3792	GGCCTGATCGGATCAAGGCTCTCTACGTAAACACCTTGCGCGGTTGACAGCGCTGGCCT	3851
Q	y	3789	-----cacgcgcgacgcgaaccatcatgaccaccaacttacocatcgaccgcg	3837
D	b	3852	GCGCGTGCTGCGAAGTAAGAGCGAGAGATATAGCTTACCAACTGACCTTTAGAACCCC	3911
Q	y	3838	tggcgctcacgatcgaggtcgcgaaggacagacgatgctcgatgcgcgctgcgcgacg	3897
D	b	3912	TGGGCGCCTCGATTGAAGTGGAGGAAGGCCAGACCCTGCTGGACGCGCGCTCGGCCAGG	3971
Q	y	3898	gcatctaatctcgcgaacgcgtgtcgcggcgctgcgcggcacctgcgaaggtgcgcgctgc	3957
D	b	3972	GCATTTACATTCCTCATGCTTCGGCCACGGCGCTGTGCGGCGACCTCTCAAGATCCAGGTTT	4031
Q	y	3958	tcgacgcgcgagaccgatccgcgatgcgaacccgttcgcgtgatggatttcgagcgcg	4017
D	b	4032	GGCAGGCGGATGTGGACCAAGCGCGAGCCAAATCCCTTCGCGCTGTATGGACATGGAGCGCG	4091
Q	y	4018	aggaaggcaagcgctcgcgtgtcgcgcacgctgcagccgcgaacccgtgatcgagggcg	4077
D	b	4092	AGGACGGCATGACGTGGCTGCTGCGCCACGCTGCAGTCCGACGCCACCATCGAGGCCG	4151

[illegible]

QY	5143	ggcagctgcacactcgaagtggccggcggtcgcagggaagcggcttctt-----ctcgggca	5197
DB	5229	GAGGCCGTGCGCATTTGCGGTGGCAGCGCCTGCACAAACCATTTTCTGCGGGCGGCC	5288
QY	5198	tggcatgtgcgggcaagcgccgcatcaacaagtaaacaggaggagagatcaccatgggtgt	5257
DB	5289	TCCCGCGGCCCAACAACCCAGCGCACACCAACAAGCAACTAAGGAGACAAACATGGGTGT	5348
QY	5258	gatcgctattggtcatgtcagctcgtgaagtgatggacatggaagcgcgctgcgtcatta	5317
DB	5349	AATGGCAATCGCCACGCCCACTTGGCGGTGATGGACATGGATGCTGGCTCAAGCACTA	5408
QY	5318	cgtacgctgtcgtcgcatcgaggaacgatgcgcgacgcgcgggcgcaactctacatgaa	5377
DB	5409	CGAAACCTGCTGGCATGAAGTCACGCTGCGCGATGTCAGCGGCAATGCTACCTCAA	5468
QY	5378	atgctggagcaaatgggacaaagtattcgtgatcctgtcgcgttcgatacggcgggct	5437
DB	5469	ATGTTGGGACGAGTGGGACAAGTACTCCTCATCTCACGCAAAAGTATCGCGCGCGCCT	5528
QY	5438	caagcatcgccctacaaaggtcgcagcagcgcgcatctcgtatgctgcgagcagcgt	5497
DB	5529	CAATCATGTGGCTTACAGGTGCAAAAGATGCCGACCTCGACAGCCTCGAGGCACGCAT	5588
QY	5498	cgaagcgtacggggtcgcgcagagatgctgccgaagcgcgctgcgcggcgctcgccg	5557
DB	5589	CGAGCCTGGGGTGTGAAGACCACCATGCTGCCGAAGGCACGACCGCTGGGCGC	5648
QY	5558	ccaactgggttctcgtcgcagcggccatgaactcgcgtgttcgcgaagaagcgct	5617
DB	5649	CATGCTGCAGTTCGATCTGCCCTAGCGGCATGAGATCGCTGTACGCCAAGAAGAGTG	5708
QY	5618	ggtggcaccgcggtcggctgcgtgaaccccgatccgtggccgacacatcccgcgctc	5677
DB	5709	CGTGGGCACGACGTGGGCTCTCTGAACCCCGACCCCTTGGCCCGATGGTCTCAAGGGCGC	5768
QY	5678	ggcgtgcactggctgcaccactcctgctgtagtgcgaactgaacccgagggcgcgct	5737
DB	5769	AGGTGCGCACTGGATAGACCACCTGCCTGCTGATGCGAGATGAATCCGAGGGCGGTAT	5828
QY	5738	gaaccgctcgagagagaacacgcgtctcattggcgcagatgtcgaactccatctggcga	5797
DB	5829	CAACACCGTGGAGGACAAACACCCGTTTCATGGCAGAGTGCATGGACTTCTTCCTGACCGA	5888
QY	5798	gcagtgatgctcgcccgcggaacacgat	5827
DB	5889	GCAGATTCGTGGCTGGCCCCCGGAAGGCAATAT	5918
RESULT	2		
AB024741			
LOCUS	AB024741	7823 bp	DNA
DEFINITION	Comamonas testosteroni phcR, phcK, phcL, phcM, phcN, phcO, phcP and phcQ genes for regulator, phenol hydroxylase subunits and ferredoxin-like protein, complete cds.		
ACCESSION	AB024741		
VERSION	AB024741.1	GI:6505654	
KEYWORDS	ferredoxin-like protein; phenol hydroxylase subunit; positive regulator of phenol-degradative genes.		
SOURCE	Comamonas testosteroni (strain:R5)	DNA.	
ORGANISM	Comamonas testosteroni		
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.		
AUTHORS	1 (sites)		
TITLE	Teramoto,M., Futamata,H., Harayama,S. and Watanabe,K. Characterization of a high-affinity phenol hydroxylase from Comamonas testosteroni R5 by gene cloning, and expression in Pseudomonas aeruginosa PAO1c		
JOURNAL	Mol. Gen. Genet. 262 (3), 552-558 (1999)		
MEDLINE	2005/761		
REFERENCE	2 (bases 1 to 7823)		
AUTHORS	Teramoto,M.		
TITLE	Direct Submission		





















Qy 2710 agaagcgcgctgttcaagatctcgcgctacgagcattcgagccgaaggct 2769  
Db 2962 GCAAGCGCGCGCTGTTCAGGATCTGGCGCTATGGATCTGGTACGCGCGGGCT 3021  
Qy 2770 ggcggagcctgcgaaggcaagatcacatcagccaccaggcgctgtcagcttctacg 2829  
Db 3022 GGAAGGACGCTGCGAGGCAAGGACCAACATCAAGCCACAGGATGGGCCACGTTCTACA 3081  
Qy 2830 gcttcaacgcgctcgcgcttccacacctgggtgcgacccgaagaaatggctggc 2889  
Db 3082 ACTAC-GGGCGCGCGCGCTTCCACAGTGTTCGAGC--GCGAAGATGCAATGGC 3138  
Qy 2890 tgtcggaagtatcccgactctgtcagcgtctactaccgcccgcgttgcactcaggg 2949  
Db 3139 TCTCGGAGAGTACCCGGAGTCTTCACACGCTGTATCGCCCGCCCTTGAATTCGGG 3198  
Qy 2950 gcgagcggcgggcgccggaacgcttctacatgaagacgctccgctgtgtgcaga 3009  
Db 3199 CGCGAGGGCGGAGGAGGACGCGCTTCTACAAAGAGCGCTGCCGATGCTGTCACGA 3258  
Qy 3010 cgtgccagatcccgatctgttaccgagccgggcaaccgcaagatcgggcgcg 3069  
Db 3259 CCTCCAGGTGCCGATGTCTTCAACGAAACGGGCGGACGCCAACAGATGCCCTATCGCG 3318  
Qy 3070 aatcgaaactcctcgcaacaagtccactcttcagcgaccactgcaaggacatctcg 3129  
Db 3319 AGACGACTACTTCGGCAACAAAGTACCACCTTCTCTCGACCACTGCAAAACATCTTCG 3378  
Qy 3130 atcagcgcgcagaaatacgtgcagcggtggtcgcgggtgcacagatccatcaggca 3189  
Db 3379 ACGATGAACCGAGAAATATGTGCAGAGCTGGGTGCGGTGCACACATCTATCAGGGCC 3438  
Qy 3190 actgctccgcgcgcgaatgcggagccggcgaggcttcgatccgctcgccgcgtgc 3249  
Db 3439 ATTGCTTCCCGGAGGTACCGACCGGACACCGGAAGCTTTCGATCCGTTGATGCCCTGC 3498  
Qy 3250 tgcactactcgcggtgacatggcgcgcaacaacctcgatttcgacggtcgcgaagacc 3309  
Db 3499 TCAAGTACTACGAGATGAACATCGGAGGGGACACTTCGACTTCGAAGGCTCGGAAAGCC 3558  
Qy 3310 agaagaacttcgcggtggtgcggccagccagccagcgaactgacctgacctgcaacgaagca 3369  
Db 3559 AGAAGAACTTTGCCGCTTGGACAGGAGAAACAGC----- 3592  
Qy 3370 atcttgacgagggccgcgaagcgccgatgcgcgaacgcggcgacaggaagacaacat 3429  
Db 3593 -----ATGAGCGTCTG 3603  
Qy 3430 ggcgctcatcgcgctcaaacctcactgacttcccggtgaagatgcgctcagagaagttcc 3489  
Db 3604 GACCAACCCGNAAGCCGCGCTTACAAAGTTTCCGATGAAGGATGTGGCGGAGAACTTTC 3663  
Qy 3490 ggcgctgctctctacgtgtgtggtggaaaccatctgatgttcccgcgccggttctgcct 3549  
Db 3664 CGCGGCGTCTGTACATCGGTGGGAAGACCACTTGATGTCTTCGCGCGCGGTCTGCCT 3723  
Qy 3550 gcgctgcgcgcgcgaatgcgcttcgcgcgctgcgcggagagtgctgcgcgccgtctta 3609  
Db 3724 GCCGCTGCCGCCACATGCGGTTTGGCGGCT-GATGACAGGTGTGCTGTGGCATCTA 3782  
Qy 3610 cgggtacaccccgacttcgcaagatcgcactggatcgcgctcagtggttcccggtcg 3669  
Db 3783 CGCGGACATCCGACTTCGAGCGTGTGACTGGAATGCGGCGCAGTGGTTCAAGTCCGG 3842  
Qy 3670 cgaagccttggcgccgga-cccgcgaagagccctggccggcaacggcctcgccgcaag 3728  
Db 3843 TCAGCCCTTGGACGCCGATGGCGGCCAAGTCCGTGGCGGAGAACGGTCTGGGCCACAGG 3902  
Qy 3729 acctgatacgttccgaccccgcttcgacgcttcgcgcgcgagcttctgcagcgc 3788  
Db 3903 ACGTATCCGTTTTCGCACCGGGGGCTTACCGGCATCAAGGGTTCGTGAGTTAAGC-- 3960  
Qy 3789 cagcgagcagcgaaccatcatgagccaaacttaccatcgagccgctggcgctcagc 3848

Db 3961 -----CATGTGCAATCAACTCACCATCGAGCCCTCGGCCGACG 4000  
Qy 3849 atcgaggtcgaaggagacagacgatctcgatgcccgctgcgccagggcatctaatc 3908  
Db 4001 CTGACGTCGGCCGAGCCAGACCTCTGTCAGCCCGCTTGCCTCGGGGTATACATC 4060  
Qy 3909 ccgcacgctgctcagcggtgtgcgacactgcgaaggtcgcgctgcgcagcgag 3968  
Db 4061 CCGATCGTGGCGCCACAGACTGTGGGTACGTGCAAGGTGCAGGTACACAGCGCGGAA 4120  
Qy 3969 accgatccccgcatgcgaaccgctt--cgcgctgattgatttcgagc-gcgaggaaggc 4025  
Db 4121 GTGGATCAGCGTCAGCAACCCCTTGGCGGCTCATGTGATATCGAGCGGGAAGAGGC 4180  
Qy 4026 aagcgctcgtgtcgcgcgaacgctcagggccgacaccgtgatcgcagggccgacgtcgac 4085  
Db 4181 AAGACGTCGGCTGTGCCACGCGCTCAGCGATGTCTGCATCGAAGCGCATGTGGAC 4240  
Qy 4086 gaggagcggatgcg---gaaatcatcccggtcaaggacttcgcgccgacgctcaacgc 4142  
Db 4241 GACGAGCGGATGCGCGCGCTCATTCCTGTCGGACTTGGTGGGCGCACGGTCACGCG 4300  
Qy 4143 atcgaaagctcacgcgcgaaccatcaagtgcgctcgaagctgtcgcaagcgcgcgc 4202  
Db 4301 ATCGACAGCTCAGCCGCCACCATCAAGGGCTGCGCTGAAGCTCGATCAGCCGATCGAT 4360  
Qy 4203 ttcagcgggccagtagctgcagctcagattccccggcctcgggcagagcgcgcttc 4262  
Db 4361 TTCCAGCGCGCCAGTAGTACGTGATGGTGAGATCCGGGCTGGGGCAACGCGCGCCTTT 4420  
Qy 4263 tgcagcgaaacgcgcgcgcgacgtgcggccacccgagagatcgaactgaacgtcg 4322  
Db 4421 TCGATGCAAA---TGCCGAGCGACGGTGGACCGCTGCGGTGAGATCGAACTTCGAGGTACG 4477  
Qy 4323 caggtgcggcggtcgcgcgcggtctacctgcagcgaactggcgagcgcgagcgc 4382  
Db 4478 CGGTACCCGCGCGCGCACCGCGGTGCTGCACGAGCAGCTTGGCGGTGGTGACAGC 4537  
Qy 4383 gtgcgctgttcggcccgctacgcccgttcttcgctcgtcgtcgccgcgcgcgatg 4442  
Db 4538 CTGCACGTGACGGGCGGTACGGCGGTCTTCTGTGCGGAGCTGCGCGCCCAAGCCCATG 4597  
Qy 4443 atcttcagccgcgggtcgggctgcagcccgctcgtatgcgcgacgtgc 4502  
Db 4598 TTGTTCTGTCGGCGGTGCGGCTGTGAGCCCGCGCTCGATGATCTCGATCTGCTG 4657  
Qy 4503 gcaagcggtcaccgcgcgcgatacagctgtctacggtcagcgacgcgcgcgagagctc 4562  
Db 4658 GAGCAGGGCTGCAGTTGCCATCACGCTCGTCAATGGCCAGCGCAGCGGAAAGAGCTG 4717  
Qy 4563 tactacacgaagaattcccgcgctgcgcgaacgcatccgaacttcacgtgcgc 4622  
Db 4718 TATGGCCACCGGAGTTTGTGCGCTGCGGATGCTCACCGAACTTCAGCTACGTCCG 4777  
Qy 4623 gcgctgcccaagcgccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4682  
Db 4778 GTGCTGTAAACGAGCCGAGGAGCGCTTGGACGGTGGCGCGCTTGTGCGACGAT 4837  
Qy 4683 gtgcgaagcgacatttcggcgacttctccgggcccagcgcgctacacctgtcgggcg 4742  
Db 4838 GCGGCCAAGCGCATTTCAACGGGACGTTTGGGGGACAAAGGCGTATCTGTGGGGCCG 4897  
Qy 4743 ccgcgagatgcagcgcgtgcatacagcagctgatgcagggcgccctgttcgagcgagc 4802  
Db 4898 CGCCGATGTGAGGTGCGCTCATGCGCGCTCATGAGGGCGGGCTGTTTCGAGAGGAC 4957  
Qy 4803 atctacagagaagtctatctcgcgcgcgagcgcaacagacgcgcgcgcgcgcgcgc 4862  
Db 4958 ATCTACACCGAGAAATTCCTCTCGGCCGCGGATGCCAACCGCGGCGGCGCTGCTC 5017  
Qy 4863 cggcggtgtga 4874  
|||||

Db 5018 AAGCGGTTTGA 5029

# RESULT 5

AF026065

LOCUS

DEFINITION

AF026065 9511 bp DNA BCT 20-AUG-1998  
Ralstonia sp. E2 positive phenol-degradative gene regulator (poXR),  
phenol hydroxylase components (poxA, poxB, poxC, poxD, poxE, poxF),  
and ferredoxin-like protein (poxG) genes, complete cds.

ACCESSION

AF026065

VERSION

AF026065.1

SOURCE

KEYWORDS

ORGANISM

Ralstonia sp. E2.

Ralstonia sp. E2

Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

Ralstonia.

REFERENCE

AUTHORS

Hino, S., Watanabe, K. and Takahashi, N.

TITLE

Phenol hydroxylase cloned from Ralstonia eutropha strain E2

JOURNAL

Microbiology 144 (Pt 7), 1765-1772 (1998)

MEDLINE

98361023

REFERENCE

AUTHORS

Watanabe, K. and Hino, S.

TITLE

Direct Submission

JOURNAL

Submitted (18-SEP-1997) Kamaishi Laboratory, Marine Biotechnology

Institute, Heita 3-75-1, Kamaishi, Iwate 026, Japan

FEATURES

Location/Qualifiers

1..9511

/organism="Ralstonia sp. E2"

/strain="E2"

/db\_xref="taxon:67939"

2014..3720

/gene="poXR"

2014..3720

/gene="poXR"

/note="PoxR"

/codon\_start=1

/transl\_table=11

/product="positive phenol-degradative gene regulator"

/protein\_id="AAC32451.1"

/db\_xref="GI:344531"

/translation="MSSSTDNFSATMRDGLSNLARRLFAMKEGSIWLGEQRMILLHT

AALGARKEFLVTGLMERAGLFMRGPHSGVGRDAELAKTMRSGHSDFGMLEGPCIH

TIEGVVVTPLVDINIAAGVYHGFELWEDSEGDVHRMGVQAQVPCWMIQIATG

YTSALMGKTLIRELECVCGGPHPKILGKPLEQWEDGEALALYOPDPVIDTILALQ

SEVEQLRALQRANDPADLVGPGFRAWNLQRAAGSSVTLLIGFVGKGRFAQ

ALHVSARAGKPFVAVNCAAIPELIESLFGVERGAFTGAHQSRAGRPERAHGGTLF

LDLGLSASQSKLLRVLOQEVERVGGNEARKYDVRVLAATNVDLAEAARQGTFRK

DLYRLNVPVTIPLRLRLDRIILAEFRTRYGARHGKKILGITDRALAEELRYDW

PGNVRLENVIERGVLANGQISAEQLFLPGAEPAPVADTTPRLGEGALPLGLEA

AVHGLDYNAGLALGEVESVLMERAMORADGNLSQAARLLGLTRPQLAYRWKTRGT

RG"

4490..4714

/gene="poxA"

4490..4714

/gene="poxA"

/note="PoxA"

/codon\_start=1

/protein\_id="AAC32453.1"

/db\_xref="GI:3445533"

/translation="MQVDIKTQIOIPLRQTYGHVARRFGDKPASRYQEAITYDQSEVN

FHYRPTWAPQFELYDKRTTAIEADMYALKDQRYYYGAYVGTGRQQAEEKFAFY

ERKGLQALPLEWRRLTDLGLRHRVMAANNFYCADYGMGTAITQACTYCAMDR

LGIAYQLSRLGLLDCNTGAALERARAAWLESEAWOPLRRFVEHSFVLTDFWTFVTQ

NIALDGLVPLVYOHADAALVACGTGLAVVTEFMDWRDEHVRVDAVVOFAAASE

ANRALLSWAGEARAQAARALRPVAAILLGGEGQAIALCLQFDARLAKLGVA"

5813..6085

/gene="poxC"

5813..6085

/gene="poxC"

/note="PoxC"

/codon\_start=1

/transl\_table=11

/product="phenol hydroxylase component"

/protein\_id="AAC32454.1"

/db\_xref="GI:3445534"

/translation="MSANVYIALQNNDDTPRIETAEANPLAVYSQFPAMVKIDAPG

RLTIVRELVDKLRDWDQLQEIHLNLSLSGNIDETDEAFTLHWSA"

6153..7667

/gene="poxD"

6153..7667

/gene="poxD"

/note="PoxD"

/codon\_start=1

/transl\_table=11

/product="phenol hydroxylase component"

/protein\_id="AAC32455.1"

/db\_xref="GI:3445535"

/translation="MDARKKLNLREKYADMTRDLGWETTYEPMDKVFPDPKYEIGIKH

DMDKWEDPFRMTMDAYWKYQSEKELVAILDSFYQNGHLNLSVDAVYLNALRLFTG

VTPLEYAHRGYAHLGRHGRGAGARVAQMQSIDELRHAQTQLHLSVYKNKTFHGFGE

WRMHDRVWYLSVPKSPEDASAGFFETITAFISFSEYFVYVLLNLLFMPMSGAAYNGD

MATVTFGSAQSDERHMTLGLVVKFLCEQDPDNIPLLQKWLDFWNGFRLLTLVG

MMMYMLPRVMSAEAWEMYPEQAGGALFKOLERYGLRLPKRYHDVATKTKRITHEA

WGTFYNAAGAFHTWI PKPDMAWLDKEY POTFARYYKPRLDYQWEROQAGERYVNG

TLPMLCOTCQIPMVESEDDPTQTCYRESSYHGKMFHFCSDGCKDIFDGEPEKYAQAW

LPVHQIYQNGCGGASLEDVLKWLRLNLGADNLDFEGSQDKNNNAWKVPHPA"

7004..8063

/gene="poxE"

7704..8063

/gene="poxE"

/note="PoxE"

/codon\_start=1

/transl\_table=11

/product="phenol hydroxylase component"

/protein\_id="AAC32456.1"

/db\_xref="GI:3445536"

/translation="MSVASIGEVREPEADREAVFHGNRLLYVGMQHLHLCAPHGCFPF

PRAMRREVEQVLFVGYGHPDFGRIDMSRVWNLGGRPMQPDLDATLENGLGHKE

VIRFRTPGLDGGGSAS"

8089..9159

/gene="poxF"

8089..9159

/gene="poxF"

/note="PoxF"

/codon\_start=1

/transl\_table=11

/product="phenol hydroxylase component"

/protein\_id="AAC32457.1"

/db\_xref="GI:3445537"

/translation="MISLTIEPIGQTIPIAQQTVDACLRSQVWLPFHACCHGLCATC

KQVYVEGVEDQGEASFAIMDFRNGQCLACCAQSDLTTEADIEDADALGLPLA

DYEAEDVDVRLTPTTRIGTLWRKGGAKVAFQGYLNLQVPCQCPQAFSLANAPDE

ELVELHVRKVEGGQATGYLHERLVAGDALRFSAPYGRFRVRSQAQMLFLAGSGLS

SPRAMIRELLAAGETLPTLVQAGARNRAELYGEDEFERALAERHPFRYVPALSDPAD

SAWQARGFAHEALAAALYADGGRADFRGHKAYLCGPPMIETACITRLMQGRLEFADI

HTEKFLSASDAQNSARSLFKI"

gene

CDS

































QY 3631 gaagatcagactgggagtcgctgagtggttcctggtcggcgagcgcgtggcgccgaacc 3690  
Db 3685 GAGGATCGATTTACGCCAGCGCTCTGGCAGCTGAACACGACCGCTTACCCCGGACTA 3744  
QY 3691 ggcgaagagcctggccgcaacagcgcctcgggcaacaagagacatgatacagttccgcaacc 3750  
Db 3745 CCGCCGCCAGCCTGGAAGCAACAGGATCGACCAAAAAGCATGCTGCGTCTGAACACGCC 3804  
QY 3751 cggcctcagagcctcggcgcgagctcttgagccgcacgcgagcagagaaacatca 3810  
Db 3805 GGGCCTGAACGGCATCCAGGGTTCCTGACGTACAGGTGTG-----TCA 3849  
QY 3811 tgaagcacaacttaccatcagcagcgcgtggggtcacagatagagtcagagaaagacaga 3870  
Db 3850 TGACTTACAACGCTACCATCGAGCCTACCGGTGAATCATCGAGTTCGAGAGGGCCAGA 3909  
QY 3871 cgaatcgaatgcgcgtgcgcgaagggcatctaatccgaacgcgtgctgtcaacggc 3930  
Db 3910 CCATCTGCAAGCGCCTTGGGCCAGGGCTGTGCTGCCATTCGCCCTGCGGCCATGGTA 3969  
QY 3931 tgtcggcacctgaagctcgcgtgctcgacggcgagacagcagatcccgcgatcgaaacc 3990  
Db 3970 CCGTGTGCACCTGCAAGGTGAGGTAGTTCGAAGGGAGGGCCACACGGCCGCCCTCAC 4029  
QY 3991 cgttcgcgctgatgatttcgagcgcgaggaagcaagcgcgtcgcgtgcgcgacgc 4050  
Db 4030 CCTTTGCCCTGATGACATGAGCGTGACGAGGGCAAGTCTTGCCCTGCTGCGGCCATGC 4089  
QY 4051 tgcagggcgacacogtgcagggccgagcgtcgacgagggagccgagatcggaatcatcc 4110  
Db 4090 CCATGAGCGATATGTTAGAGCGGATATCGACGTGATCCGGATTTTCGCCGCCATC 4149  
QY 4111 cggtcaggaacttcggcgagtcagcgatcagcgatgaacagatcagcgccagcatcaagt 4170  
Db 4150 ACGTGCAGGACTACCGCGGGTGTGTCAGCGCCCTGGTTCGACCTGTCGCCGACCACTAAG 4209  
QY 4171 cgaatcgcgtgaagctgcgcagcgcgtccgttcacggcgccgagtaagtcagctcgc 4230  
Db 4210 GTGTGCACATCAAGCTCGATCGGCTATGACCTTCCAGCGCGGGCAATACATCAACTGA 4269  
QY 4231 agattcccgctcggcgagcgcgctgttcctcgatcgagaaacgcgcgcgcagctgc 4290  
Db 4270 CCGTCCCGGGATCGAAGGCTCCGCGCTTCTCCCTGGCCAAACCGCGCCAGG--- 4326  
QY 4291 cggccacggcgagatcgaaactgaacgtgcgcgaggtgcggcgcgccgcgcgcgcgc 4350  
Db 4327 -----CCGACGAGTGGAGTTGATATCCCGCTGGTCGAGGGCGGCGCGCCACCGGCT 4380  
QY 4351 acctgacgagcaactggcgagcgagcgcgctgcgcgtcggcgccgtaacgcgcgcgc 4410  
Db 4381 TCATCCATCGGCNACTGAAGGTTCGGCGATCGGTCGAGCTGTCCGGCCCTACGGGCAGT 4440  
QY 4411 tcttgctgctgctcggcgcgcgccgagtgatcttcaatggcgggcggttcggggctgt 4470  
Db 4441 TCTTGTGCTGGTTCGAGCGCGCGACCTGATCTTTCATCGCGCGGCGCTGTGGCCTGT 4500  
QY 4471 cgaagcgcgctgatgatcgagacatgctcgcaagcgcggtcacgcgcgcgcgcgcgcgc 4530  
Db 4501 CCAGCCCGCATGCTGATGTTCTTCAGCTGCTGGCGCAGGGCGACCGCGCAGATCACCC 4560  
QY 4531 tggctacggtcagcgagcgcgagagctctactacacgagcaaatcccgcgctgg 4590  
Db 4561 TGTTCAGGGCGCGGCNAACCGCGCGAGCTTTACACCGCGGAGCTGTTTGAAGAACTGG 4620  
QY 4591 ccgaacgccatccgaacttcacgtacgtgcggcgctgtccgaagggcgacccgcgcgcgc 4650  
Db 4621 CCGCGCGCCACTCAACCTTACGTACGTGCGCGCACTCAACACGAGCGCCACGACGCCGG 4680  
QY 4651 gcggcgacgtcgcaaggggttcgtgacagctcgcgaaagggacacatctcgcgcgac 4710  
Db 4681 AGTGGCAGGGCTTCAAGGCTTTGTCCACGACGCGCGCCAGGCCACTTTCACAGGTGCTT 4740  
QY 4711 tctccgggcaaccagcgctacatgtgcgggpcgcccgcgcgcgcgcgcgcgcgcgcgcgc 4770

Db 4741 TCAGTGGCCACAAGCCCTATCTCTGCGTCCGCGCGATGATGACGGCCATCACCA 4800  
QY 4771 cgtcgaatcagggggcgccgtgttcgagcgcgcacatctatcaagagaagtctatctcggcgg 4830  
Db 4801 CCTGTACAGGGTGGCTGTTCGAGCGGGACATTTTATGAGGCGCTTCTTACCGCGC 4860  
QY 4831 ccgacgcgaacagacgcgcagcc-----cgtgttcgcgggtgtgacatg-gagcg 4883  
Db 4861 CCGATGGGGCTGAGGACAGCACCCGCTTCGGCTTGTTCAGCGCATCTGAGTGGCTCAT 4920  
QY 4884 gggccgcgctatgcgggagcgtcacgacgcgcagaccagcagcgtatcgtcgtgtc 4943  
Db 4921 GAGCAGCCCCCGTTTACGGTGCAGAAACCAACAGCGCCAGTCTTTACTGCGCGCC 4980  
QY 4944 cggcgagtcgctgctggcgccatgcgcaactcggcgcgccgacatcccggtcggtg 5003  
Db 4981 GGATCAGTCCGTCTGCGGGCATGGAGGAGCAGGGAAGCGTTTGCCTGCGGTCGGTTG 5040  
QY 5004 cctgaacggcggtgcgcgtgtcgaagtcgcgcgtgctgcgcgtgcggtgcgcgaact 5063  
Db 5041 CCGTGGTGGTGGCTGTGGGTGTGCAAGGTGCGGTGCTCAGCGG---CGACTACCACTG 5097  
QY 5064 cggcgccgctcagcgcgtgcccgtgagcgcggaagaagaacacgacgctcagcgctgc 5123  
Db 5098 CGGCAGGATGAGTGCAGTCAAGTTCACCGGAGCGCGCAGCAGGCGCTTGGCTTGGC 5157  
QY 5124 gtccgcgctcgtccgggagcgcgtcgaactcgaagtggccgcccgcgcgcgcgcgcgc 5183  
Db 5158 CTGTCAACTGATCCACGCGTGTCTGTACATCGAA----- 5194  
QY 5184 gttcttcggcgatggcgtgc 5243  
Db 5195 ---TCCCTTCCCAAGTGCAGAAACAATCCCTGACAACAAGAAACATTAAGAGGTGTCGTC 5251  
QY 5244 ctccacctgggtgtgctgctgattggtcgtcgtgaggtgagtgatgagacatggaagcg 5303  
Db 5252 ATGAANAAGAGTATGCGCCCTGGACACCTCCAGCTTCGCGTGTGAATCGAGTCT 5311  
QY 5304 gcgtgcgcttattacgtacgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5363  
Db 5312 CCGTAAAGCAGTATCCCGGACCTGCTCGGCTGATCGAGAGGATCGCGAGCAGCAGGA 5371  
QY 5364 aactctacctgaatcgtggagcaatggagcaatgagcaatgctgctgctgcgcgcgcgcgc 5423  
Db 5372 CGTATCTACCTGAAGGCTGGAGCGAGTGGACAAAGTTCCTCCGTGGTCTACGCGAGGCC 5431  
QY 5424 gatcaggcgggctcgaagcatgcgcgcctacaaggtcgagcagcgcgcgcgcgcgcgcgcgc 5483  
Db 5432 GCCGAGCGCGGCATGGATTTCATGGCTTCAAGGTGCTCGAGCAGCAGCACCCTGAGCAGC 5491  
QY 5484 ctgcagcagcgcatcgaagcgtacgggacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5543  
Db 5492 CTCACCGACGACCTGATCCCGTACGGCTGCTGGTTCGAGCTGATGCGCGCGCGGAGCTC 5551  
QY 5544 ccggcgctcggcgcccaactcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5603  
Db 5552 AAGACTGCGTGGCGGGTGGCTTCCAGGCGCTTCCCGGGCACCTTTCGAGCTCTCAT 5611  
QY 5604 gcgaagaaggcgctgggtggcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5660  
Db 5612 GCGGACAAGGAGTACACCGGCAATGGGCTTGGCGAGGTCAACCCGAGGCGCTTGGCGC 5671  
QY 5661 gacgaatcccggtcgc 5720  
Db 5672 CGCAACCTCAAGGATCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5731  
QY 5721 aaccgggagcgcggtgaaacgc 5774  
Db 5732 CTGCAAGCCACTATGAGCTTTACCGAGGTGCTCGGTTTCTACCTGGCGCAG 5785









[illegible][illegible]













Db	61	GGTGCTCGATGCGTTACGCGAGAAACACGCGTTCTCTCGCGTGACCGACGCGCGCTACAT	120	LOCUS	DB5083	5552 bp	DNA	BCF	11-NOV-1997
Qy	2090	caacgcctgaagctgtttctccaggcggtgacccgctgaataacctcgccaccgcggg	2149	DEFINITION	Acinetobacter sp. gene for DMS oxygenase components, complete cds.				
Db	121	CAACGCGCTGAAGCTGTTTCATCCAGGGCGGTACGCCGCTTGAATATCTCGCGCATCGCGG	180	ACCESSION	DB5083				
Qy	2150	cttcgcgcattgcgcggcgacttcaccgcggagggcgcgcatcgctgccagatgca	2209	VERSION	DB5083.1	GI:2605607			
Db	181	TTTCGCGCACGCTCGCGCGCACTTCACCGGGGAGGCGCGCATCGCTCGCGATGCA	240	KEYWORDS	DMS oxygenase component.				
Qy	2210	gtcgatcgacgagctgcggcgactaccagaccgaaacgcatgctgcgacgtacacaa	2269	SOURCE	Acinetobacter sp. (strain:20B) DNA.				
Db	241	GTCGATCGACGAGTTCGGGCACTACACAGCCGAAACGACGCGATGTCGACGTACACAA	300	ORGANISM	Acinetobacter sp.				
Qy	2270	gtttctcaacgggttcacatcactcgaaaccagtggttcgacgcgctggtacctgtcgt	2329	REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;				
Db	301	GTTCCTTCAACGGCTTCATCATCTCGAACCACTGGTTCGACCGCGTCTGGTACCTGTCGGT	360	AUTHORS	Moraxellaceae; Acinetobacter.				
Qy	2330	gcccgaagtgtttctcaggagcggtattcgtcggggcccgttcagttcctgaccgcggt	2389	TITLE	1 (bases 1 to 5552)				
Db	361	GCCGAAGTCGTTCGAGGACGGTATTCGTCGGGGCGGTTCCGAGTTCCTGACGGCGGT	420	JOURNAL	Omori,T.				
Qy	2390	cagttctcgttcgaatactgctgtaagaaacctgctgttcgctgcgcttcgctggtgcgc	2449	MEDLINE	Submitted (07-MAY-1996) to the DBJ/EMBL/GenBank databases. Toshio				
Db	421	CAGCTTCTCGTTCGAATACGTGCTGACGAACCTGCTGTCGCGCTTCATGTCGGGGCG	480	FEATURES	Omori, The University of Tokyo, Biotechnology Research Center;				
Qy	2450	cgctcaacaggtgacatgtgcacggtcaagttcggcttctccgcgagtcggacgaatc	2509	source	Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan				
Db	481	CGCGTACAACGGCGACATGTCCACCGTACACGTCGCGCTTTTCCGCGCAGTCGGACGAATC	540	gene	(E-mail:asegyo@nongo.ecc.u-tokyo.ac.jp, Tel:03-3812-2111(ex.3067),				
Qy	2510	gcgtcacatgacgtcgcgcagtcgaatgcatcaagtctcgtcgcgaacagggacccggacaa	2569	CDS	Fax:03-5802-3326)				
Db	541	CGCTCACATGACGCTCGGCATCGAGTGCATCAAGTTCCTGCTCGACGAGGCCGACAA	600		2 (sites)				
Qy	2570	cgtcccgatcgtgacgcgtggatcgacaagtgttctcgtgcgcggtaccgcgtcgtgac	2629		Horinouchi,M., Kasuga,K., Nojiri,H., Yamane,H. and Omori,T.				
Db	601	CGTCCCATCGTGCAGCGCTGGATCGACAAGTGGTCTTCGGCGCGCTACCGGCTGCTCAC	660		Cloning and characterization of genes encoding an enzyme which				
Qy	2630	gctggtcgcatgatgatggactacatgcagcccaagcgctgatgactgacgcgcaatc	2689		oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B				
Db	561	GCTGGTCGGATGATGATGGACTACATGACCGCAAGCGCGGTGATGAGCTGGCGCGAGTC	720		FEMS Microbiol. Lett. 155 (1), 99-105 (1997)				
Qy	2690	gtggagatgtacgcgagcagagaacgcgcgcgtgttcgaagatctcgcgcgctacgg	2749		98005684				
Db	721	GTGGGAGATGTACCGGGAGCAGAACGGGCGCGCTGTTCAAGGATCTCGCGCGCTACGG	780		Location/Qualifiers				
Qy	2750	cattcgagccgaaggctggcgagcgctgcgaaggcgaagatcacatcagccacca	2809		1. 5552	/organism="Acinetobacter sp."			
Db	781	CATTCCGAGCGGAAGGGCTGGCAGGACGCCCTGCGAAGGCAAGGACCAACATCAGCCACCA	840		/strain="20B"	/db_xref="taxon:472"			
Qy	2810	ggcgtggtcagcttctacggttcaacgcggcctcgcgctgcacacctgggtgcgcac	2869		/db_xref="GI:2605608"	/gene="dsob"			
Db	841	GGCATGTGCACATCTACGGCTTTCACGCGGGCTGCGCGCTTCCACACGTCGGGTGCCGAC	900		/transl_table="1"	/gene="dsob"			
Qy	2870	cgaagacgaatggctggtcgtcgcgcgaagtatcccgactcgttcgacccgctactaccg	929		/product="DMS oxygenase component"	/protein_id="BAA23331.1"			
Db	901	CGAGGACGAGTGGCTGGCTGTCGCGAAGTATCCCGACTCGTTCGACCGCTATTACCG	960		/transl_table="1"	/db_xref="GI:2605609"			
Qy	2930	ccgcgcttcgatcactcgtggcgagcgagggccggcgccggcgacacctctctacatgaagac	989		/transl_table="1"	/gene="dsoc"			
Db	961	CCCCGCTTCGACCTACTCGGGGAGAGCAGCGGCGGCAAGCGCGCAACCGCTTCTCATGAAGAC	1020		/codon_start=1	/gene="dsoc"			
Qy	2990	gctccgatgctgcagacgctccagatcccgatcgtgttaccggagcgggcgcaacc	3049		/product="DMS oxygenase component"	/protein_id="BAA23332.1"			
Db	1021	GCTCCCATGCTGTGCGCAGACTGCCAGATCCCGATGCTGTTCACCGGACCGCGGCAACCC	1080		/db_xref="GI:2605610"	/transl_table="1"			
Qy	3050	gacgaagatcgcgcgcgcaatcgaaactacctcggc	3086		/transl_table="1"	/transl_table="1"			
Db	1081	GACGAAGATCGCGCGCGGAGTCGAACCTACCTCGGC	1117		/transl_table="1"	/transl_table="1"			
RESULT 13									
DB5083									









```

/gene="ORF5"/
/codon_start=1
/transl_table=1
/product="phenolphydroxylase component"
/protein_id="CAA85384.1"
/db_xref="GI:535284"
/db_xref="SPTREMBL:O43982"
/translation="MTVQAIKERYQEPELDLOQNYGENMLLFITGWDHHTLFCSAHAFLY
VSPKSQLIDIQIOAGFGPHDFKHIDMSKVEFRINRLLOADPFSKLEIDLGFDRK
SLRFFVTPLAGYQGTHV"
3792..4844
/gene="ORF6"
/codon_start=1
/transl_table=1
/product="phenolphydroxylase component"
/protein_id="CAA85385.1"
/db_xref="GI:535285"
/db_xref="SPTREMBL:O43983"
/translation="MSYQVTEIPAGTIIQVEEDOTILDAALRGVWLPFACGHGTGCT
CKVQTDFDYDVGCEAPFDALDERENKLACCPESDMVTEADVEDDEDLGLVLV
ODYQAQVETLDSPIIKGVRLDRPMOFQAGQYINQLPNIEGTRAFSIANTPSDK
NLIHLRIHKVGGGAATRYHDELVSUEEMALSQGYQFFVRKSDQNVIYIAGSGSL
SLQSMILDLLEHGDIIRYLFOGARDMAELNREKFQIVKYPNFYIPALNAPKPE
DOWTGTGYVHEAVANYFENKSCGHAYLCGPPMPMDAAISTLMQSLRFKEKDITHREF
LSAADGANGQSREFGE"
4973..5890
/gene="ORF7"
/codon_start=1
/transl_table=1
/product="catechol 1,2-dioxygenase"
/protein_id="CAA85386.1"
/db_xref="GI:535286"
/db_xref="SPTREMBL:O43984"
/translation="MMNNRQITDSLVOQMNVATATGEVNLRVQOIIVVRLIGDLFQAI
E DLNNSQTELWKGLELYLTDAGQANELGLAAGLGEHYLDLRADAACKAGITGGTPRT
IGNEPLXVAGAPESVPGARMDGSESHVDALIIEGNVDTTAGOIIPNAKVEIWHANS
INTSFDPKSQSANLRSIFTFDQQQYIAQTNPVGYCGGPBEGTTQALLNLGRHNCR
PSHVHFVSPAGYFKLUFTQNIISGDXYLMDDFAFATRDGLIALDVLTDLAKIKQYNL
NKAFKHIFNFQIVQADQVPLQRLLWE"
BASE COUNT      1816 a   1199 c   1243 g   1647 t
ORIGIN
Query Match          13.4%; Score 779.2; DB 1; Length 5905;
Best Local Similarity 50.6%; Pred. No. 1.6e-71;
Matches 2239; Conservative .0; Mismatches 2093; Indels 94; Gaps 11;

Qy 452 aggaagcaagaagtaccatcgactgaaagacagctgcacatacgaacccgtcggcacacct 511
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 437 AGGAGAAGCTAATGACTTTAGAAAATTAAAACTAACCATCCAATTTGCCAACCCGATTTCGCAGACTT 496

Qy 512 ttgcgatgtcgcgacaacatcgcgcgacgaacgagcgagcgagcgctaccaggaaagcca 571
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 ATGCGTATATCG---AACGCGTTTTTGAGCGCAAGCCGCCAACACGCTATCAAGAAGTCA 553

Qy 572 tgatggcgcgagccccagagagaacttcattaccggccgagacctgggccggactacy 631
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 GCTTTGATATTCAAAGCTTCACAAACTTTCACTATCGCCCCCTTTGTGAAGCCACACAAA 613

Qy 632 agatcttcgactcgctcgcgctcgcgatccgagtgaggaaactggttacggttaagatc 691
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 CCTTAATGATAAAACCCTATACCCGCTTTGCAAAATGAAGATGTGATGCAATTTAAAGACC 673

Qy 692 cgcgccagttctactacgcgctgtggcgacacacgcggcgcgccagcagagatgcgatgg 751
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 674 CTCGCCAATTTATTACGGTCTTATGTGCACAATCGTGCACGCTCTCGAAGATACCCTGTG 733

Qy 752 agtgcgaactcgadttogtogaactcgcgccggatgatcggcctgatcgcgacgacgtgg 811
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```





Qy 2570 cgtgccgatcgtcagcgctggatcgacaaagtgggttcttgccgagcgtaccggctgctgac 2629  
Db 601 TGTGCCCATCGTCAGCGCTGGATCGCAAAATGGTTCGTGGCGGGCTACCGCGTCTGAC 660  
Qy 2630 gctggctcgatgatgatggactacatcgacccaaagcgcgtgatgagctggcgcgagtc 2689  
Db 661 CCTGTGGCCATGATGCAAGGATTACATGCTGCCAAAGCGCGTATGAGCTGGAAGGAAGC 720  
Qy 2690 gtggagaatgtacccagcagaacgagcgcgcgctgttcaagatctcgcgcgctacgg 2749  
Db 721 CTGGAGAGATGATGCCGAGGAAACGGCGCGCGCTCTGTTCGCGACCTGGCGCGCTACGG 780  
Qy 2750 cattcgcagccgaagggtcgtgcaggacgcctgcgaagcgaagatcacatcagccacca 2809  
Db 781 CATTCGCGAGCCGCGAGCTGGAAGCTGGCCGTGCAAGGTAAGGACACATCAGCCACCA 840  
Qy 2810 ggcgtgtcgcagcttctacgcttcaacgcgcctcgcgcgttcacacacctgggtgcgcgac 2869  
Db 841 GGCCTGGAACATCTCTACAACATACGGCGCGCGCACCTTTTCACACCTGGGTGCCCGG 900  
Qy 2870 cgaagacgaatggcgtgctgctggcggaagtatcccgactcgttcgacgcgtactaccg 2929  
Db 901 CGAGCAGGAGATGAGTGGCTGTCCGAAATAACCCGACAGCTTTGACAAAGCACTACCG 960  
Qy 2930 cccgcgcttcgactcactgggggcgagcagcccgagggccgggcaacgccttctacatgaagac 2989  
Db 961 TCCCGCGCTGGAGCACTACCTTTGCCGAGCAGCAGCGCGGCAAGCGCTTTTACAGCAAGAC 1020  
Qy 2990 gctgccgatcgtgtgccagacgtgccagatcccgcgtgttccaccgagccgggcaaccc 3049  
Db 1021 CCTGCCCATGCTGTGCACCACTGCTCCAGATCCCTATGGGCTTACCGAGCCCGGTGACGC 1080  
Qy 3050 gacgaagatcgcgcgcggaatcgaactactcgtggcaacaagtccacttctgcagcga 3109  
Db 1081 CACCAAGATCTGCTATCGCGAGTCCGACTAGAGGGCAGCAAAATACCACTTCTGCAAGCA 1140  
Qy 3110 ccaactgcaagga 3121  
Db 1141 CGGCTGCANAGCA 1152



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2000, 15:50:55 ; Search time 153.15 Seconds  
(without alignments)  
9520.858 Million cell updates/sec

Title: US-09-430-029-1  
Perfect score: 5828  
Sequence: 1 gatcatttcacaaatgcgc.....tcggcccggaacacgacgac 5828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues  
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	2490.6	42.7	9785	1 T44457	DNA encoding tolue
2	2072.8	35.6	12019	1 X18867	Alcaligenes sp. Po
3	1854.8	31.8	4800	1 X05350	Phenol hydroxylase
4	1854.8	31.8	4800	1 X01590	Pseudomonas putida
5	327.2	5.6	652	1 X04453	Strain M07 oxygena
6	201	3.4	6379	1 O79569	Nocardia corallina
7	201	3.4	6379	1 T17418	Nocardia corallina
8	137.2	2.4	9785	1 T44457	DNA encoding tolue
9	109.8	1.9	1240	1 N60800	Portion of plasmid
10	109.8	1.9	1992	1 N60855	Sequence encoding
11	109.8	1.9	2004	1 N60911	Sequence encoding
12	109.6	1.9	924	1 Q04017	Metapyrocatechase
13	100.2	1.7	958	1 N30167	Sequence of gene x
14	94.4	1.6	114955	1 X53491	Human adenosine A1
15	88.6	1.5	114955	1 X53491	Human adenosine A1
16	73.8	1.3	1266	1 X25773	S. erythraea erythr
17	73.8	1.3	3412	1 X25772	S. erythraea erythr
18	72.4	1.2	44377	1 T78508	Platenolide synth
19	72.4	1.2	44377	1 T80414	Sugar biosynthesis
20	72.2	1.2	3756	1 T72684	Amicolatopsis medi
21	71	1.2	53789	1 V21187	eryA region of S.
22	69.4	1.2	29879	1 Q46806	Amicolatopsis medi
23	68.4	1.2	5676	1 V21186	Cephalosporin anti
24	61	1.0	23666	1 Q10190	Actinomadura hibis
25	60.8	1.0	8169	1 V26609	Gtfid gene from Amy
26	60.6	1.0	1224	1 V60558	Streptomyces venez
27	59.8	1.0	15872	1 T68715	Microspora therm
28	59.4	1.0	2634	1 V22334	Total DNA sequence
29	59.2	1.0	30001	1 T61016	S. aureofaciens DN
30	59.2	1.0	30001	1 X05110	Orf virus genomic
31	59	1.0	4020	1 T91361	Soil derived pepti
32	58.8	1.0	1204	1 V69560	Pentaclethra macro
33	58.2	1.0	1227	1 V84121	

ALIGNMENTS

RESULT 1

T44457	ID	T44457 standard; DNA; 9785 BP.			
AC	T44457:				
DT	29-JAN-1997	(first entry)			
DE	DNA encoding toluene ortho-monoxygenase.				
KW	tom; pTOM; self-transmissible; constitutive; bioindicator; pollutant;				
KW	breakdown; trichloroethylene; TCE; degradation; ss.				
OS	Pseudomonas cepacia strain PRL-23.				
FH	Key	Location/Qualifiers			
FT	CDS	2647..2757			
FT		/*tag= a			
FT		/product= tomA0			
FT		/note= "no start or stop codons included; encodes			
FT		W06805"			
FT	CDS	2872..3866			
FT		/*tag= b			
FT		/product= tomA1			
FT		/note= "encodes W06800"			
FT	Misc_difference	3862..3866			
FT		/*tag= c			
FT		/note= "apparent deletion of nucleotide; encodes			
FT		cysteine and stop"			
FT	CDS	3900..4175			
FT		/*tag= d			
FT		/product= tomA2			
FT		/transl_except= (4173..4175, aa:Glx)			
FT		/note= "encodes W06801"			
FT	CDS	4198..5753			
FT		/*tag= e			
FT		/product= tomA3			
FT		/note= "encodes W06802"			
FT	Misc_difference	5116..5117			
FT		/*tag= f			
FT		/transl_except= (5116..5117, aa:Ala)			
FT		/note= "apparent deletion of nucleotide"			
FT	Misc_difference	5133..5135			
FT		/*tag= g			
FT		/note= "encodes Ala-Arg; apparent deletion of 3			
FT		nucleotides"			
FT	CDS	5750..6030			
FT		/*tag= h			
FT		/product= tomA4			
FT	Misc_difference	6026..6027			
FT		/*tag= i			
FT		/transl_except= (6026..6027, aa:Ala)			
FT		/note= "apparent deletion of nucleotide; encodes			
FT		W06803"			
FT	CDS	6185..7245			
FT		/*tag= j			
FT		/product= tomA5			
FT		/note= "encodes W06804"			
FT	Misc_difference	6224..6225			
FT		/*tag= k			
FT		/note= "apparent insertion"			
FT	Misc_difference	6397			

Infected cell prot  
The nucleotide seq  
Sorghum bicolor (L  
Streptomyces prist  
GtfC gene from Amy  
snBR gene encoding  
S. clavuligerus cl  
Tyactone synthase  
Sequence encoding  
Neomycin phosphor  
Maize nitrite redu  
Hybrid srmg/tylg O



[illegible][illegible]















```

Db 3017 AGCTACACGCGATCGCGTTCCTCCACTTCTGCTCGAGCGCTGCAAGGACATCTTCGATGGC 3076
Qy 3135 gagccgcgaataacgtgcagcgctggtcgctgcggtgcaccagatccatcagggcaactgc 3194
Db 3077 GAGCGCGAAGAGTAGCGCGCAGCGCTGGTGTGCTGTGTCACACAGATCTACGACGGCAACTGC 3136
Qy 3195 ttcccgcccgatgcggaaccggcgcgagggttcgactccgctgcgcgcgcggtgctgcac 3254
Db 3137 -----GGCGGTGATCGCTGGACGAGCTGCTCAAG 3166
Qy 3255 tactacgcggtgacgatggcgcgacaaacctcgatttcgacgctcggaagaccagaag 3314
Db 3167 TGGTAGCCGATCAACCTTGGCGCGGACAACTGGATTGGAAGTTTCGAGGATCAGAA 3226
Qy 3315 aactcgcggtggtgcggtgcgagcccaacgcgcaactgaccgcgcaacgaagaatctt 3374
Db 3227 AACTGGAAAGCGCTGGAAGGGGTGCC----- 3252
Qy 3375 gacgagggccgcgaagcgcgatgcgcgaacgcggtgcgacgagagagacaacatggcg 3434
Db 3253 -----GGGACAGCGTGCCTGAGCGCGCGCATACAAAGAGATAAGGAGACCAACCATGTCCG 3307
Qy 3435 tcatcgctcaaaacctacgacttcccggtgaaggtgcggtgcgagagccggtgtccacggca 3367
Db 3308 TCGTATCCATTGGCCCTTACAGCTTCGAGCGCGCGCATGCGGAGGCGCGTGTTCACGGCA 3307
Qy 3492 cgcgcgtgcttaegtgtgtggaaacacatctgatttccccggcgcgcttctgcctgc 3551
Db 3368 ACCGCTGCTGTATATCGGCTGGGACCGCCACCTGCTGCTTTGTGCGCGCATTCGCCCTC 3427
Qy 3552 cgtgcgcgcgcgaatcgcttgcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3611
Db 3428 CTTGCGCGCGCTGATGCGCTGCGGATGCTGCGGAGAACGTCGCTGCGGCGGTCTACG 3487
Qy 3612 gctatcaccccgacttcgcgaagatcgactggatcgctgcgagtggttccggtgcggcg 3671
Db 3488 GCTATCACCCCGACTTTGCCCGCATCGACTGGAGCGCGTGGAGTGGCTGCGCGCGCGCG 3547
Qy 3672 agcgtggcgccgagaccgcgaagagcctggccgcgcgaacgcgcgcgcgcgcgcgcgcgc 3731
Db 3548 AGCCCTGGCAGCCAGACTTACCGCACGCTGGAAGAGAACGCGCTGGGCGCCACAAAGCGG 3607
Qy 3732 tgatcagcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3791
Db 3608 TGATCCGCTTAGCACACCGGGCTGGATGGCAITGGCGGAGTGGCGAGCTGAACAGTGGC 3667
Qy 3792 gcgacgagcgaaccatatgagccaccaacttaacctgacgcgcgcgcgcgcgcgcgcgcgc 3851
Db 3668 AGCTAAGGTAGGAGACACAAGATGTATTCCTGACCATTTGAACCGATCGGGCAGACCATC 3727
Qy 3852 gaggtcgaggagacagacgatgctcgtatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3911
Db 3728 CCCATCGCGCGCGGCGCAGACCGCTGTGATGCTGCTGCCCTGCGCACGCGGTGTGGCTGCCG 3787
Qy 3912 cagcgtgctgcagcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3971
Db 3788 CACGCTGCTGCCAGCGGCTGTGCCACCTGCAAGGTGCGAGTGGTGGAGGGCGAAATTC 3847
Qy 3972 gatcccgcgatgcggaacccgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4031
Db 3848 GAGCATGGAGAGGCTCCAGCTTCGCGCTGATGGACTTCGAGCGCGACAGCGGGCAGTGC 3907
Qy 4032 ctgcgctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4091
Db 3908 CTGGCTTGTCTGCGGCACCGCGCAGTGCACATGTTGATCGAGGCGCGATATCAGGAAGAC 3967
Qy 4092 ccgcatgcggaatacatctcccggtcaaggacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4151
Db 3968 GCCACTGCTGCGGCTTCCGCTGCTGACTATGCTGCGCGAGTGGTGGAGCGCGCGCG 4027
Qy 4152 ctacgcgcgacat-----caagtcgactccgctgaagctgtgcgacgcgcatcgccttc 4205
```

```

Db 4028 CTGACCCCCACCATCCGCGGCATCTGGCTGCGCGTGAAGGGGGGGCGCGCTCCCTTC 4087
Qy 4206 caggcggccagtagtcagctcagctcagatctcccgctcgggcagagcgcgcgcttcctcg 4265
Db 4088 CAGCGCGGCCAGTACTCAACCTGCGCTGCGGGCTGCGACGACCGCGTGCCTTCG 4147
Qy 4266 atcgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4325
Db 4148 CTGGCAACCGTCCCGCGGATGACCTGG-----TGGAGCTGATGTGCGGCGG 4195
Qy 4326 gtgcggcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4385
Db 4196 GTGGAAGCGGCGGACGCTACCGGCTACCTGCACGATACGCTGCTGCTGGTGGTGACAACTC 4255
Qy 4386 cgcctgtcggcccgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4445
Db 4256 GGGTTTTCGCGGCTTACGCGCGCTTCTGCTGCGCAAGTCAGCGCAAAAGCCGATGTG 4315
Qy 4446 ttcatggcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4505
Db 4316 TTCTTGGCGGCGGCTCGGGCTTGTCCAGCCCGCGGCATGATCTTGACATGCTGGCT 4375
Qy 4506 agcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4565
Db 4376 GCCGCGAGACCTTCCGATCAGCTGCTGCAAGCGCGCGCAACCGACGAGCTGTAC 4435
Qy 4566 taccacgacgaattccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4625
Db 4436 TACGACGAGGGCTTCGCTGCTGCTGCGCGCGCGCGCACCTCCCGTATGTGCGCGCG 4495
Qy 4626 ctgtccgaagcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4685
Db 4496 CTCCTCGAGAACCGCGCGGACAGCGGCTGGGACGCGCGCGCGGCTATGTGCTGACGTC 4555
Qy 4686 gcgaaggcacatttcgg-----cggcgacttctcggcaccacgcgcgcgcgcgcgcgcgcgc 4733
Db 4556 CTGCACGCGCTTTACGCCAATGGCGCGACCGCGCGCTTCCGTGGCCACAAAGSCCTATCTG 4615
Qy 4734 tgcggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4793
Db 4616 TCGCGCGCGCGCGCGGATGATCGAAGCCTGCATCCGACAGTTGATGACGGCGCGGTTC 4675
Qy 4794 gagcgcgcacatcatcagagaagttcatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4850
Db 4676 GAGGAGGACATCCACCGAGAAATTATCTCGGCGCGCGGACGACAGACAGCGCGCGC 4735
Qy 4851 agccgcgcttc 4862
Db 4736 AGCCGCGCTTC 4747

RESULT 4
X01590
ID X01590 standard; DNA; 4800 BP.
AC X01590;
DT 05-MAY-1999 (first entry)
DE Pseudomonas putida phe operon.
KW Chlorinated ethylene decomposition; phe2; pheA; pheB; pheC; pheD; pheE;
KW Chlorinated ethylene-decomposing gene; ds.
OS Pseudomonas putida.
FH Key Location/Qualifiers
FT CDS 127..345
FT FT /*tag= a
FT FT /product= phe2
FT FT 434..1429
FT FT /*tag= b
FT FT /product= pheA
FT FT 1440..1712
FT FT /*tag= c
FT FT /product= pheB
FT FT 1754..3268
FT FT /*tag= d
FT FT /product= pheC
```



Db 2057 GCGGTGACGCCGTGGAGTAGCGGGCGCACCGGGCTATGCCACCTCGGGCGCACTTT 2116  
QY 2175 acgggagggcgccgcatcggtgccagatgcagtcgatcgacagcgtcgccgcactac 2234  
Db 2117 CGGGGGCGCGCGGGGTGGGGCGGACAGTGCAGTCGATCGACGAGCTGCGCCACGCC 2176  
QY 2235 cagaccgaacacgatcgatgtcgagctacaaagttcttcaacgggttccatcactcg 2294  
Db 2177 CAGACCCAGATGCACAGCGCTGTGGTCTACACAAGTACTTCAACGGCTTCGGCGAATGG 2236  
QY 2295 aaccagtggttcacccgctgtgtacctgtcggtgcgaagtcgttcttcagaggcgcg 2354  
Db 2237 CAGCACATGCACACCGGGTGTGGTACTCTCGTGGCCCAAGTCTATTTCGAAGACGCC 2296  
QY 2355 tattcgtcgggcggttcgagattcctgacccggttcagcttctcgttcgtaatacgtgctg 2414  
Db 2297 ATGAGCGGGGGCGGTTCGAGTTTCATACCGCGATCTCTCTCTTCCTTCAGTATGTGCTG 2356  
QY 2415 acgaacctgctgttcgtgcccgttcattcggtcgggcgccctacacaggtgcacatgctgac 2474  
Db 2357 ACCAACCTGTGTTCATGCCCTTCATGTGCGGTGCGGCTTACAAACGGCGACATGGCCACG 2416  
QY 2475 gtcacgttcggtctctcgcgcagtggtggaacgaatcgcgttcacatgacgtcgcgcacatg 2534  
Db 2417 GTGACCTTCGGCTTCGCGCACAGTCGAGCAATCGCGCATATGACGCTGGGGCTGGAA 2476  
QY 2535 tgcatcaagttcctcgtcgaacgagaccggaacaaogtgcgcagtcgtgcagcgtcgatc 2594  
Db 2477 GTGGTGAAGTTCTGTGCCAGCAGGACCCGGCAATATCCCATCTCTGTCAGAAAGTGGCTG 2536  
QY 2595 gacaagtgttcgcgcggtaccggctcgtcagcgttcgagtcggttcgagatgatggaactac 2654  
Db 2537 GACAAATGGTTCTGGCGCGGCTTTCGGCTGCTTACGCTGTGTCGGCATGATGATGGAGTAC 2596  
QY 2655 atcgagcccaagcgcgtgtagtgcgcgagtcgtggagatgtacgcgcgagcagaac 2714  
Db 2597 ATGCTGCCCAAGCGCGTGTGCTGGCGGCAAGCCCTGGGAGATGTACTTCGAGCAGGCC 2656  
QY 2715 ggcgcgctgttcgaaggtatctcgcgctacggtcgtcaggttcgagcgaaggtcgag 2774  
Db 2657 GCGGGCGCGTGTTCAGGACCTTGGAGCGCTAGCGCTGCGGCTGCGCATGCCCAAGTACCACGAG 2716  
QY 2775 gacgctcgaaagcagaatcatcatcagccaccagcggtggttcgagttctcaggttc 2834  
Db 2717 GTGGCCACCAAGAACAGGACCGCATCCACACGAGCGCTTGGGCGACCTTCTCAACATAC 2776  
QY 2835 aacgcgctcgtcggttccacacctggttcgacgaagaaacgaacaaatgggtcggtcgtcg 2894  
Db 2777 GCTGCGCGCGCGCTTCATACCTGGTGCCTCCCAAGCAGGACGAGATGGCTGGCTGACG 2836  
QY 2895 gcgaagtatccccgactggttcgaccgtactacccgctcgttcgactcgttcgactggtcgag 2954  
Db 2837 GAGAAATATCCCGACACTTTCGGCGGCTACTACAAGCCGCGCTCGATTTACTGGCAGGAA 2896  
QY 2955 caggccagggcgcaaccgttctcatatgaagcgtgcgagtcggttcgttcgcaagcgtgc 3014  
Db 2897 CGCCAGCAGCAAGCGAGCGCTTCTACACGCCACCGCTGCGGATGTGTGCCAGACCTGC 2956  
QY 3015 cagatcccgatgctgttcacgcagccgggcaaacccgacgaagatcgcgcgcgcgcaatcg 3074  
Db 2957 CAGATCCCGATGTGTCTCGGAGCGGACGACCCACCCACGACCTTCTACCGGGGAGAGC 3016  
QY 3075 aactacctcggaacaaagttccacttctgagcgcagaccactgcaaggacatcttcgatcac 3134  
Db 3017 AGCTACACCGCATCGCGTTTCCACTTCTGTCTCGACGCGCTGCAAGGACATCTTCGATGGC 3076  
QY 3135 gacccgcagaaaatcagtcgaagcggtggttcggttcgaccacatccatcaaggaactgc 3194  
Db 3077 GAGCCCGCAGAAGTACGCGCAGGCGCTGGTGGCTGTGCACAGATCTTACCAGGCAACTGC 3136  
QY 3195 ttccccgcgcgatcggaaccggcgcgagggttcgatccgctcgcgcgcggtgctcgac 3254

Db 3137 -----GGCGGTGATCGCTGGACGAGCTGCTCAAG 3166  
QY 3255 tactacggtgacgatggccgcgacaaactcgatttcgacggtcgggaagaccagaag 3314  
Db 3167 TGGTACCGCATCAACCCCTGGCGGACCAACTGATTTGAAGGTTCCGAGGATCAGAAG 3226  
QY 3315 aactcgcggtggtggtcggtccgagccacgcgcgaactgacccgcgaacgaagaactctt 3374  
Db 3227 ACTTGAACCCCTGGAAGGGGTGCC----- 3352  
QY 3375 gaagagggccgcggaagcgcgatgcygcaacgcyggccgacagagacaaaacatggtccg 3434  
Db 3253 -----GGGCACGGCTGCTGAGCGCGCGCATACAAGAGATAAGGAGACCACTGTCG 3307  
QY 3435 tcacggtcgtcaaaacctacgacttccggtgaaggtgcggtcgcgaagaagtcttcgg--- 3491  
Db 3308 TCGTATCATTTGGCCCCCTACACGTTTCGAGCCCGCCGATCGGGAGCGCGTCTTCACGGCA 3367  
QY 3492 cgcgctgctctacgtgtgctgggaaaaacatctgatttcccggtccggttctgctgc 3551  
Db 3368 ACCGCTGCTCTATATCGGCTGGAGCCGCCACCTGCTGTGTTTGTGCGCCGCAATTGCCCTTC 3427  
QY 3552 cgtgcgcgcgcacatgccgttcggcgctggtgcgcgcgagctgctgcgcgcgtctacg 3611  
Db 3428 CCTTGGCGCGTCTGATCGCTGCGCATGTCGTGGAGAACGTCGCTGCCGGCGCTCTACG 3487  
QY 3612 gctatacccgacttcggaagatcgactgggagtcgctgcagatggttccggtcggggcg 3671  
Db 3488 GCTATCACCCCGACTTTTGGCCGCTGCTGCTGAGCCGCGTGGAGTGTGCTGCGGCGCGC 3547  
QY 3672 agcgtggtgcgcgacccgcggaagagcctgcccgcgaacgctcggccgcaacaagacc 3731  
Db 3548 AGCCCTGGCAGCCACTTTCACCGCAGCTTGAAGAGAACGGCTTGGGCCACCAANGCGG 3607  
QY 3732 tgatcaacttccgc 3791  
Db 3608 TGATCCGGTTACGCACACACCGGGCTGGATGGCATTTGGCGCAGTGGCAGCTAACAGTGGC 3667  
QY 3792 gcggagcgaacacatcatgagccacaaacttaccatcagcgcgcgcgcgcgcgcgcgcgc 3851  
Db 3668 AGCTAAGGTAGGAGACACAAGATGTATTCTCTGACCATTTGAACCGATCGGGCAGACCATC 3727  
QY 3852 gaggtcgaagaagacacgacgatcgatcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3911  
Db 3728 CCATCGCGCGCGCGCAGACCGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3787  
QY 3912 cagcgtgctgcacgggtggtgcgcgcacgtcgaagtcgcgcgcgcgcgcgcgcgcgcgcgc 3971  
Db 3788 CACGCTGCTGCCACGGGCTGTGCGCCACCTGCAAGTGCAGGTGCTGAGGGCGGAAATTC 3847  
QY 3972 gatccgcgcgatcgaaacccgttcgctgctgatgatttcgagcgcgcgcgcgcgcgcgcgc 4031  
Db 3848 GAGCATGGAGAGGCTCTCAGCTTGCCTGCTGATGACTTCGAGCTTCGAGCGCAGACGGCAGTGC 3907  
QY 4032 ctgcgctgctgc 4091  
Db 3908 CTGCTGCTGCGCCACCGCGCAGTCCGACATGCTGATGCTGAGGCGCGGCTATTCAGGAAGAC 3967  
QY 4092 ccggatcgcgaaatcatcccggtcagggacttcgcccgcgcgcgcgcgcgcgcgcgcgcgc 4151  
Db 3968 GCGGACTGCTGCGCTGCGCTGGTGGTACTATCGTCTCCGAGGTGGTGGAGGCCCGCGCG 4027  
QY 4152 ctcaagccgacct-----caagtcgatccgctgaagctgttcgagcgcgcgcgcgcgcgcgc 4205  
Db 4028 CTGACCCCACTCCCGCGCATCTGGCTGCGCTGAGGGGGCGCGCGCGGCTGCCCTTC 4087  
QY 4206 caggcgggcagtcagtcgagatcccggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4265  
Db 4088 CAGCGCGCGCGAGTACCTCAACCTGCGCGTGCCTGGGCTGCGCACCGCGCGGCTGCTTCG 4147  
QY 4266 atcggaacgc 4325  
Db 4148 CTGCCCAACCTCCCGCGCATGACCTCG-----TGGAGCTGCATGTGCGCGCG 4195











CC tom. The Tom gene is present on a large self-transmissible plasmid  
 CC denoted pTOM. The enzyme is capable of degrading trichloroethylene  
 CC (TCE), a hazardous pollutant. The pTOM plasmid is transmissible and  
 CC expressible in other bacteria, thus many bacteria can be genetically  
 CC altered to constitutively degrade TCE, esp. in bioreactors or  
 CC TCE-contaminated environments. P. cepacia PRI-23 contg. pTOM does not  
 CC need exogenous chemical inducers and is capable of functioning under a  
 CC diverse set of conditions. Also it does not require an inducer that is a  
 CC co-substrate for the enzyme required to break down TCE, in effect, it is  
 CC not subject to competitive inhibition.  
 SQ Sequence 9785 BP; 2225 A; 2697 C; 2815 G; 2048 T;

Query Match 2.4%; Score 137.2; DB 1; Length 9785;  
 Best Local Similarity 70.1%; Pred. No. 1.4e-15;  
 Matches 213; Conservative 0; Mismatches 88; Indels 3; Gaps 2;

QY 5524 tgcgtccggaagcgctgcgcgagcattccggtccgacgaactgcggttcctcgtgcgcgagcg 5583  
 DB TCCTGCTCTTCGGTACCTTCCTTCGACCGTCTGATGCTGCAGTTCAACCTGCCGAGCG 8915  
 QY 5584 gcatgaactgcggtgtctgcgaagaagcgctggtggccaccgcggtgcgctgta 5643  
 DB GACATGAGATGCGGCTCTTTTCGACGAGGAATATGTGGCACCAGGAGTCCGAAACGACCA 8855  
 QY 5644 acccgcgtgcggtgcgcgagcattccggtccggtgcggtgcggtgcggtgcggtgcg 5703  
 DB ATCTGACCGGTGGCCGATGACGTCAAGGGGGCAG--GGCATGCTGCTGATCAGTCC 8797  
 QY 5704 tgcgtatgtgcgaactgaacccgagcggtgcggtgcggtgcggtgcggtgcggtgcg 5763  
 DB TATTGATGTGTGAATCAATCCGGA-CACGGCGTGAATCGGGTGGCGGAAATACCCGAT 8738  
 QY 5764 tcatgcccaggtgtctgcgaattccatctgcccagagcggtgaggtgcggtgcggtgc 5823  
 DB TCATGAAGGAATGCTGCACTTTTATCTGCTGACGAAGTATGTAGTCCGATAGTA 8678  
 QY 5824 cgat 5827  
 DB 8677 GCAT 8674

## RESULT 9

N60800  
 ID N60800 standard; DNA; 1240 BP.  
 AC N60800;  
 DT 20-JUN-1991 (first entry)  
 DE Portion of plasmid pSLMK1 encoding metapyrocatechase.  
 KW E.coli; ds.  
 OS Bos taurus.  
 PN J60241899-A.  
 PD 30-NOV-1985.  
 PF 15-MAY-1984; 097106.  
 PR 15-MAY-1984; JP-097106.  
 PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 PA (CENG ) CENTRAL GLASS KK.  
 PA (HODO ) HODOGAYA CHEM IND KK.  
 PA (NIPS ) NIPPON SODA KK.  
 DR WPI; 86-018108/03.  
 PT Cyclic double-stranded DNA - for the prodn. of  
 PT meta-pyro:catechase.  
 PS Claim 9; Page 524; 19pp; Japanese.  
 CC The plasmid allows for the expression of metapyrocatechase under the  
 CC control of a promoter/operator from the tot plasmid of Pseudomonas  
 CC putida, in an E.coli expression system with an ampicillin resistance  
 CC marker. The 5'-terminal overhangs it's complement by AATT- the 3'-  
 CC terminal is overhang by -TTAA.  
 SQ Sequence 1240 BP; 272 A; 358 C; 357 G; 253 T;

Query Match 1.9%; Score 109.8; DB 1; Length 1240;  
 Best Local Similarity 50.3%; Pred. No. 6.2e-11;  
 Matches 298; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

QY 5184 gtcttctggtgcatggtgcgcgcatggtgcgcgcatggtgcgcgcatggtgcgcgcatggtgcgcg 5243  
 DB GTATCTTCGGCCACGTTGGCGGAAACAAACCTGACAACTATGAAGATGTGACGTC 145  
 QY 5244 ctaccattggtgtgcatggtggttcattggttcattggttcattggttcattggttcattggttc 5303  
 DB ATGACAAAGGTGTAAATGCGACCGGCGCATGTGACGTGCTGCTACTTGGACATGAGCAAG 205  
 QY 5304 gcgtgctgctcattcgtacgtgcgtgctgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgt 5363  
 DB GCCCTGGAACACTACGTCGAGTGTCTGGCCTGATCGAGATGGACCTGTGACGACGACGCGC 265  
 QY 5364 aacgtctacgtgaaatgctgggacgaatgggacgaatgggacgaatgggacgaatgggacgaatgg 5423  
 DB CGTGTCTATCTGAAGGCTTGGACCGAAGTGGATAAATTTTCCCTGGTGTACGCGAGGCT 325  
 QY 5424 gatcagcggtggtcgaagcgtgcgcgtcgaagcgtgcgcgtcgaagcgtgcgcgtcgaagcgtgc 5483  
 DB GACGAGCGGGCATGGATTTTATGGGTTTCAAGGTTGTGGATGAGGATGCTCTCCGGCAA 385  
 QY 5484 ctgcagcagcgtcgtgcgaagcgtgcgcgtgcgcgtgcgcgtgcgcgtgcgcgtgcgcgtgcgcgt 5543  
 DB CTGAGCGGGATCTGTATGGCATATGCTGCTGCCGTTGAGCAGCTACCCGAGGTGAACGT 445  
 QY 5544 ccggcggtgcgcgcgaactgcggttcctcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5603  
 DB AACAGTTGTGGCGCGCGGCTGCGCTTCCAGCCCGCTCCGCGCATCTACTTCGAGTTGTAT 505  
 QY 5604 gcgaagagcggtggtggtgc 5660  
 DB GCACACAAGGAATATCTGGAAGTGGGTTTGAATGACGTCATCCGAGGATGCGCG 565  
 QY 5661 gaagacattccgggtgc 5720  
 DB CGCGATCTGAAGGTATGCGCGCTGTGCGTTTCGACCAACGCCCTCATGTATGCGCGCAA 625  
 QY 5721 aaccggagcggtgcgaacgc 5773  
 DB TTGCGCGGACCTATGACCTGTTCCACCAAGGTGCTCGGTTTCTATCTGCGCGCA 678

## RESULT 10

N60855  
 ID N60855 standard; DNA; 1992 BP.  
 AC N60855;  
 DT 01-OCT-1991 (first entry)  
 DE Sequence encoding portion of metapyrocatechase gene and ampicillin  
 DE resistance gene.  
 KW E.coli; catechols; ds.  
 OS Pseudomonas putida mt-2.  
 PN J61067490-A.  
 PD 07-APR-1986.  
 PF 07-SEP-1984; 187792.  
 PR 07-SEP-1984; JP-187792.  
 PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 PA (CENG ) CENTRAL GLASS KK.  
 PA (HODO ) HODOGAYA CHEM IND KK.  
 PA (NIPS ) NIPPON SODA KK.  
 PA (TOYJ ) TOYO SODA MFG KK.  
 PA (NISC ) NISSAN CHEM IND KK.  
 DR WPI; 86-128931/20.  
 PT Circular double chain DNA having P1 promoter - contains  
 PT meta-pyro:catechase and operator of Lambda phage.  
 PS Claim 7; Page 477-9; 30pp; Japanese.  
 CC The sequence is a component of the synthetic circular DNA which may  
 CC transform an E.coli host for the expression of metapyrocatechase.  
 CC The metapyrocatechase gene is linked to the 3'terminal of the PL  
 CC promoter (N60853) having the amp r gene at it's own 3'terminal.  
 CC The product is useful in production of important organic  
 CC intermediates esp. catechol derivatives.  
 CC See also N60853-4.  
 SQ Sequence 1992 BP; 511 A; 529 C; 498 G; 454 T;

















GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 26, 2000, 15:21:14 : Search time 1948.25 Seconds  
(without alignments)  
13191.571 Million cell updates/sec

Title: US-09-430-029-1  
Perfect score: 5828  
Sequence: 1 gatcatttcataaatgcgc.....tcggcccgaggcaacacgac 5828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues 10495684  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

1:	em_est1.*	44:	gb_est25.*
2:	em_est2.*	45:	gb_est26.*
3:	em_est3.*	46:	gb_est27.*
4:	em_est4.*	47:	gb_est28.*
5:	em_est5.*	48:	gb_est29.*
6:	em_est6.*	49:	gb_est30.*
7:	em_est7.*	50:	gb_est31.*
8:	em_est8.*	51:	gb_est32.*
9:	em_est9.*	52:	em_est20.*
10:	em_est10.*	53:	em_est21.*
11:	em_est11.*	54:	em_est22.*
12:	em_est12.*	55:	em_est23.*
13:	em_est13.*	56:	em_est24.*
14:	em_est14.*	57:	em_est25.*
15:	em_est15.*	58:	em_est26.*
16:	em_est16.*	59:	gb_est33.*
17:	em_est17.*	60:	gb_est34.*
18:	em_est18.*	61:	gb_est35.*
19:	em_est19.*	62:	gb_est36.*
20:	gb_est1.*	63:	gb_est37.*
21:	gb_est2.*	64:	gb_est38.*
22:	gb_est3.*	65:	em_est27.*
23:	gb_est4.*	66:	em_est28.*
24:	gb_est5.*	67:	em_est29.*
25:	gb_est6.*	68:	em_est30.*
26:	gb_est7.*	69:	gb_est39.*
27:	gb_est8.*	70:	gb_est40.*
28:	gb_est9.*	71:	gb_est41.*
29:	gb_est10.*	72:	gb_est42.*
30:	gb_est11.*	73:	gb_est43.*
31:	gb_est12.*	74:	gb_est44.*
32:	gb_est13.*	75:	em_est31.*
33:	gb_est14.*	76:	em_est32.*
34:	gb_est15.*	77:	em_est33.*
35:	gb_est16.*	78:	em_est34.*
36:	gb_est17.*	79:	gb_est45.*
37:	gb_est18.*	80:	gb_est46.*
38:	gb_est19.*	81:	gb_est47.*
39:	gb_est20.*	82:	em_est35.*
40:	gb_est21.*	83:	em_est36.*
41:	gb_est22.*	84:	em_est37.*
42:	gb_est23.*	85:	gb_est48.*
43:	gb_est24.*	86:	gb_est49.*
		87:	gb_est50.*
		88:	gb_est51.*
		89:	gb_est52.*
		90:	gb_est53.*
		91:	gb_est54.*
		92:	gb_est55.*
		93:	gb_gss1.*
		94:	gb_gss2.*
		95:	gb_gss3.*
		96:	gb_gss4.*
		97:	em_gss1.*
		98:	em_gss2.*
		99:	em_gss3.*
		100:	em_gss4.*
		101:	gb_gss5.*
		102:	gb_gss6.*
		103:	gb_gss7.*
		104:	gb_gss8.*
		105:	gb_gss9.*
		106:	em_gss5.*
		107:	em_gss6.*
		108:	em_gss7.*
		109:	em_gss8.*
		110:	em_gss9.*
		111:	em_gss10.*
		112:	em_gss11.*
		113:	gb_gss10.*
		114:	gb_gss11.*
		115:	em_gss12.*
		116:	gb_gss12.*

117: gb\_gss13:\*  
118: gb\_gss14:\*  
119: gb\_gss15:\*  
120: gb\_gss16:\*  
121: gb\_gss17:\*  
122: gb\_gss18:\*  
123: gb\_gss19:\*  
124: em\_gss13:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	85.8	1.5	1618	80	AW727506	AW727506 GA_Ea001
C 2	80.4	1.4	1440	80	AW727557	AW727557 GA_Ea001
C 3	74.6	1.3	1711	80	AW731135	AW731135 GA_Ea001
C 4	73.2	1.3	925	122	CNS0091P	AL053013 Drosophil
C 5	73	1.3	925	122	CNS0091P	AL053013 Drosophil
C 6	72	1.2	935	122	CNS006XK	AL066051 Drosophil
C 7	71	1.2	1411	80	AW731189	AW731189 GA_Ea001
C 8	70.8	1.2	1201	80	AW731172	AW731172 GA_Ea001
C 9	70.4	1.2	932	122	CNS0072Q	AL066742 Drosophil
C 10	69.8	1.2	1626	80	AW731212	AW731212 GA_Ea001
C 11	69.6	1.2	1338	80	AW731160	AW731160 GA_Ea001
C 12	69.6	1.2	1501	80	AW731188	AW731188 GA_Ea001
C 13	68.8	1.2	1382	80	AW731200	AW731200 GA_Ea001
C 14	68.8	1.2	1622	80	AW731133	AW731133 GA_Ea001
C 15	68.6	1.2	1467	80	AW731219	AW731219 GA_Ea001
C 16	68	1.2	1205	117	AQ893581	AQ893581 HS_4832_A
C 17	67.2	1.2	1210	80	AW731177	AW731177 GA_Ea001
C 18	66.8	1.1	1167	80	AW731228	AW731228 GA_Ea001
C 19	66.4	1.1	1201	80	AW731172	AW731172 GA_Ea001
C 20	66.2	1.1	1307	80	AW731210	AW731210 GA_Ea001
C 21	66.2	1.1	1697	80	AW727558	AW727558 GA_Ea001
C 22	65.6	1.1	980	80	AW731225	AW731225 GA_Ea001
C 23	64.8	1.1	1468	80	AW731223	AW731223 GA_Ea001
C 24	64.6	1.1	932	122	CNS0072Q	AL066742 Drosophil
C 25	64.6	1.1	1347	80	AW731204	AW731204 GA_Ea001
C 26	64.4	1.1	1618	80	AW727506	AW727506 GA_Ea001
C 27	64	1.1	1099	80	AW731227	AW731227 GA_Ea001
C 28	64	1.1	1126	80	AW731180	AW731180 GA_Ea001
C 29	64	1.1	1415	80	AW727479	AW727479 GA_Ea001
C 30	63.8	1.1	1512	80	AW731163	AW731163 GA_Ea001
C 31	63.8	1.1	1519	80	AW731138	AW731138 GA_Ea001
C 32	63.6	1.1	1086	80	AW727532	AW727532 GA_Ea001
C 33	63.6	1.1	1243	80	AW727510	AW727510 GA_Ea001
C 34	63.2	1.1	935	122	CNS006XK	AL066051 Drosophil
C 35	62.8	1.1	1328	80	AW727531	AW727531 GA_Ea001
C 36	62.8	1.1	1723	80	AW731150	AW731150 GA_Ea001
C 37	62.6	1.1	1674	80	AW731181	AW731181 GA_Ea001
C 38	62.4	1.1	990	80	AW731206	AW731206 GA_Ea001
C 39	62.2	1.1	1138	38	AI374006	AI374006 SMOVAFAP
C 40	62	1.1	1382	80	AW731194	AW731194 GA_Ea001
C 41	61.8	1.1	1205	117	AQ893581	AQ893581 HS_4832_A
C 42	61.8	1.1	1387	80	AW731151	AW731151 GA_Ea001
C 43	61.6	1.1	1315	80	AW731169	AW731169 GA_Ea001
C 44	61.4	1.1	587	74	AW585306	AW585306 707094H01
C 45	61	1.0	1430	80	AW731156	AW731156 GA_Ea001

ALIGNMENTS

RESULT 1  
AW727506/c 1618 bp mRNA EST 20-APR-2000  
LOCUS AW727506 GA\_Ea0012K16 Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION GA\_Ea0012K16, mRNA sequence.



















```

Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemons.edu
High quality sequence stop: 1382.
Location/Qualifiers
1. .1382
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="GA_Ea0010K15"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
90 a 609 c 551 g 119 t 13 others
BASE COUNT
ORIGIN

```

	Query Match	1.23;	Score 68.8;	DB 80;	Length 1382;
	Best Local Similarity	43.7%;	Pred. No. 0.00024;		
	Matches 385;	Conservative 0;	Mismatches 493;	Indels 4;	Gaps 2;
Qy	4193	gccgatccgttccaggcgccagtagtcagctcagatcagattccggccctcgggcagag	4252		
Db	465	ccgcgcgtgtccggggcgccgctgtccggcgccgttacgcggcggttcgcgcgcgcgcgc	524		
Qy	4253	ccgcgcgttctcgtatcgcaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	4312		
Db	525	ccgttgcggcgccanngcggggncgtgtggcacccggggtgaggggtggcgagcagaccg	584		
Qy	4313	gaacgtgcgcgcaggtgccgc	4372		
Db	585	ccgc	644		
Qy	4373	gggc	4432		
Db	645	gccccgc	704		
Qy	4433	gc	4492		
Db	705	gc	764		
Qy	4493	g-gacctgtctcgaacgc	4551		
Db	765	gc	824		
Qy	4552	gc	4611		
Db	825	ccgc	884		
Qy	4612	cgactgtccgc	4668		
Db	885	ccgc	944		
Qy	4669	ggttcgtgcacgc	4728		
Db	945	cgtcgc	1004		
Qy	4729	acctgtgc	4788		
Db	1005	cgctgtccgc	1064		
Qy	4789	tgcttcgacgc	4848		
Db	1065	gcgttcgc	1124		
Qy	4849	gc	4908		
Db	1125	gc	1184		
Qy	4909	atcgc	4968		

Db	1185	CGCGCGCCCGCGCGGTCGCGGACCGCGCGCTCGCGCGCGCGCGCGCGCGGCGCN	1244
Qy	4969	gcgaaactcgccgcgcgcgcatctccgtcggtctcctgaacgcgcggtgcgcgtgtgc	5028
Db	1245	GCgtGCCCTGCCCGTCCTCCCGGTCGCGGTCCTCCGCTGCACCGCGCGCGCGCGCGCGG	1304
Qy	5029	aaggtgcgcgtgctgcgcgcggtgcgctgcgaagctcgcggccgcg	5070
Db	1305	CAGCGCGCTCGCGCGCGCGCGGAGGTGCGCGCGCGCGCG	1346
RESULT	14		
LOCUS	AW7311133/c		
DEFINITION	GA_Ea0010A01 Gossypium arboreum 7-10 dpa fiber library Gossypium	EST	20-APR-2000
ACCESSION	AW7311133	arboorem cdna clone GA_Ea0010A01, mRNA sequence.	
VERSION	AW7311133.1	GI:7628790	
KEYWORDS	EST.		
SOURCE	Gossypium arboreum.		
ORGANISM	Gossypium arboreum		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
AUTHORS	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Malvales;		
TITLE	Malvaceae; Gossypium.		
	1 (bases 1 to 1622)		
	Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.		
	An integrated analysis of the genetics, development, and evolution		
	of the cotton fiber		
JOURNAL	Unpublished (2000)		
COMMENT	On Jan 6, 2000 this sequence version replaced gi:6676789.		
	Contact: Wing RA		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel: 864 656 7288		
	Fax: 864 656 4293		
	Email: rwing@clemson.edu		
	High quality sequence stop: 1622.		
FEATURES	Location/Qualifiers		
	1..1622		

```

FEATURES
    source
        1. .1622
            Location/Qualifiers
                /organism="Gossypium arboreum"
                /strain="AKA"
                /cultivar="8400"
                /db_xref="taxon:29729"
                /clone="GA_Ea0010A01"
                /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                /tissue_type="Fibers isolated from bolls harvested 7-10
                    dpa"
                /lab_host="E. coli"
                /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
            122 a      801 c      493 g      112 t      94 others
                BASE COUNT
                ORIGIN

```

[illegible]









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:12:01 ; Search time 92.32 Seconds  
(without alignments)  
248.587 Million cell updates/sec

Title: US-09-430-029-3  
Perfect score: 1723  
Sequence: 1 MTIELKTVTDIKLRHTFAHV.....ALDEVREQFHARAARLGIAL 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues 225878

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_12.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1139	66.1	335	2 O30592	O30592 burkholderi
2	1101.5	63.9	330	2 Q92NP6	Q92NP6 comamonas t
3	1021.5	59.3	336	2 Q52570	Q52570 pseudomonas
4	832	48.3	331	2 O84959	O84959 raistonia s
5	778.5	45.2	331	2 Q52171	Q52171 pseudomonas
6	770.5	44.7	342	2 Q52162	Q52162 pseudomonas
7	640.5	37.2	333	2 Q43979	Q43979 acinetobact
8	640.5	37.2	333	2 Q32429	Q32429 acinetobact
9	222.5	12.9	330	2 O87802	O87802 pseudomonas
10	193	11.2	329	2 Q51943	Q51943 burkholderi
11	190.5	11.1	332	2 O69182	O69182 alcaligenes
12	179.5	10.4	343	2 Q53025	Q53025 nocardia co
13	167.5	9.7	341	2 Q92ET3	Q92ET3 xanthobacte
14	134	7.8	332	2 O07072	O07072 burkholderi
15	121.5	7.1	395	2 O06117	O06117 methylcyst
16	105.5	6.1	1092	3 O13462	O13462 emericella
17	105.5	6.1	1092	3 P78622	P78622 emericella
18	100.5	5.8	567	11 Q9WUL9	Q9WUL9 mus musculu
19	100.5	5.8	570	11 Q9WUM0	Q9WUM0 mus musculu

20	97.5	5.7	567	4	Q13499	Q13499 homo sapien
21	97.5	5.7	8563	2	O54297	O54297 streptomyce
22	96.5	5.6	517	2	P72449	P72449 streptomyce
23	96.5	5.6	750	12	Q9YQ39	Q9YQ39 pseudorabie
24	96	5.6	1238	2	O68032	O68032 rhodobacter
25	95	5.5	329	2	P72466	P72466 streptomyce
26	94	5.5	527	2	O9X5X9	O9X5X9 pseudomonas
27	93.5	5.4	387	2	O31029	O31029 erwinia car
28	92	5.3	687	2	O60103	O60103 xanthobacte
29	90.5	5.3	986	5	O76122	O76122 bombyx mori
30	90	5.2	539	2	Q53005	Q53005 rhodopseudo
31	89.5	5.2	415	2	O53456	O53456 mycobacteri
32	88.5	5.1	467	2	O52853	O52853 rhizobium l
33	88.5	5.1	474	1	O9YE75	O9YE75 aeropyrum p
34	88.5	5.1	509	2	O33107	O33107 mycobacteri
35	88	5.1	570	4	O15333	O15333 homo sapien
36	88	5.1	743	4	O9Y4J8	O9Y4J8 homo sapien
37	88	5.1	784	2	O86858	O86858 streptomyce
38	88	5.1	1309	10	O64428	O64428 chlorella v
39	87.5	5.1	669	2	O56151	O56151 streptomyce
40	87.5	5.1	1147	2	O87848	O87848 streptomyce
41	87	5.0	560	2	O51515	O51515 pseudomonas
42	87	5.0	560	2	O923X9	O923X9 pseudomonas
43	87	5.0	573	2	O33089	O33089 mycobacteri
44	87	5.0	1019	12	Q98630	Q98630 rice dwarf
45	87	5.0	1522	4	O15085	O15085 homo sapien

## ALIGNMENTS

RESULT 1  
O30592 PRELIMINARY: PRT: 335 AA.  
AC O30592;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JAN-1998 (TREMBLrel. 05, Last annotation update)  
DE CRPB PROTEIN.  
GN CRPB.  
OS Burkholderia pickettii (Pseudomonas pickettii).  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Ralstonia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PROJ;  
RA OLSEN R.H., KUKOR J.J., BYRNE A.M., JOHNSON G.R.;  
RL J. Ind. Microbiol. 0:0-0(1997).  
DR EMBL; AF012632; AAB67106.1; .  
SQ SEQUENCE 335 AA; 37000 MW; 103F073F CRC32;

Query Match	66.1%;	Score 1139;	DB 2;	Length 335;
Best Local Similarity	64.5%;	Pred. No. 2.6e-86;		
Matches 216;	Conservative 41;	Mismatches 74;	Indels 4;	Gaps 1;
QY 1	MTIELKTVTDIKLRHTFAHV	QAOINIGDKTATRYQEGMGAQPOENHYRTWDPDVEIFD	60	
DB 1	MNIDLQAREITLQRTFAHVARLIGEGKAATRYQEAATYQAQSMANFHYRTWDPHALEYD	60		
QY 61	PSRSATRMANWYALKDPRQFYASWATTRARQODAMESNFEEVSRMIGLMRDDVAARA	120		
DB 61	AGRSRLDLADWQALKDPRQFYATWTWTRARQAEVANEQFVESRGLASMGDALRDRA	120		
QY 121	LDVLLPLRHAAGANNNAQICALGYGTVFAPAMFAMNDNLGVAQYTLRLALAMAEPDV	180		
DB 121	LQVLLPLRHWANNNASICAYGYGTAFAPAMFAMNDNLGIAQLITRLALTLDPAV	180		
QY 181	LEAAKATWTRDAWQPLRRVVEDTLVVADPVELFIAQNLALDGLLYPLVYDVRFDRIAL	240		
DB 181	LEAGKLAWEIDPRWQGLRRYVEDSVVQDPVELFVAQNALDGLLYPLIYGHFVDDHLAL	240		
QY 241	EGGSAAVAMLTAFMPWHETESNRWDVAVVTKMA----	AESDDNRALLARWTRWSARAEAA	296	

```
Db 241 OGGTAVAILTSFPEWHDETARWIDGVIKTWANGGEPEAAGNRALLSQWFSQWADRAQAA 300
Qy 297 LAPVAARALQDAGRAALDEVEQFHARAARLGIAL 331
Db 301 LAPVAEMALGEGHAALGEVRAALDAAATKLGSL 335

RESULT 2
Q9ZNP6
ID Q9ZNP6 PRELIMINARY; PRT; 330 AA.
AC Q9ZNP6;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN APHL.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
RN [1]
RC STRAIN-TA441;
RX MEDLINE; 99018839.
RA ARAI H., AKAHIRA S., OHISHI T., MAEDA M., KUDO T.;
RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:
RT organization and regulation of the genes involved in phenol
RT degradation.";
RL Microbiology 144:2895-2903(1998).
DR EMBL; AB004679; BAA34170.1; -.
SQ SEQUENCE 330 AA; 37138 MW; 5566866E CRC32;

Query Match 63.9%; Score 1101.5; DB 2; Length 330;
Best Local Similarity 62.9%; Pred. No. 3.1e-93;
Matches 207; Conservative 43; Mismatches 78; Indels 1; Gaps 1;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYEIFD 60
Db 1 MNIELSAREIOPLRHTFARVAA-YTGDKPASRYLEATLGVPQTHFHYRPLWEPEFELED 59

Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFESVRRMIGLMDDDVAARA 120
Db 60 TGTAIRADWNLDRDPROYYATWTTTRAKQOEAMANTQFVESRNLAKMSDALRAKA 119

Qy 121 LDVLPLRHAAGANMNAQICGCTGTAPAMFHAMDNGLGVAQYLTRLALAMAPDV 180
Db 120 CGVLPURHVANGNMNNSQICSGYCTGTAPAMFHAMDNGLGVAQYLTRLGLVMDPEGV 179

Qy 181 LEAKATWTRDAAWQPLRRYVEDTLVVDPVLEFIAQNALDGLLYPLVYDRFVDERIAL 240
Db 180 LEAGKNDWMQHPAQWPLRHLVLEDLVVDQDPELFVAQNMVLDGLLYPLVYNHFVDERVT 239

Qy 241 EGSVAAMLTAFPPEWHTESNRWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 300
Db 240 OGGTAVAILTSFPEWHDETARWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 299

Qy 301 AARALQDAGRAALDEVEQFHARAARLG 329
Db 300 AELALGERGAALVSEVRQSLDIRAKAGL 328

RESULT 3
Q52570
ID Q52570 PRELIMINARY; PRT; 336 AA.
AC Q52570;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1996 (Tremblrel. 08, Last annotation update)
DE ALPHA SUBUNIT-TERMINAL OXGENASE COMPONENT.
GN TMB.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-JS150;
RX MEDLINE; 96035667.
RA JOHNSON G.R., OLSEN R.H.;
RT "Nucleotide sequence analysis of genes encoding a toluene/benzene-2-
RT monooxygenase from Pseudomonas sp. strain JS150.";
RL Appl. Environ. Microbiol. 61:3336-3346(1995).
DR EMBL; L40033; AAA88457.1; -.
SQ SEQUENCE 336 AA; 37042 MW; E4D0F654 CRC32;

Query Match 59.3%; Score 1021.5; DB 2; Length 336;
Best Local Similarity 60.7%; Pred. No. 1.2e-76;
Matches 204; Conservative 40; Mismatches 87; Indels 5; Gaps 2;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYEIFD 60
Db 1 MNIDLQAREITPLRQTFAHVARLIGEGKAATRYQEAATYARSPWRTFTIARRGTPTSCCT 60

Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFESVRRMIGLMDDDVAARA 120
Db 61 RAAAAIRLADCEALKDPQFYATWTTTRARQOEAVANFOFVESRGLASMGDALRDRA 120

Qy 121 LDVLPLRHAAGANMNAQICGCTGTAPAMFHAMDNGLGVAQYLTRLALAMAPDV 180
Db 121 LQVLLPLRHHVANGANMNVICAYGCTGTAPAMFHAMDDLGIAQLITRLALTDEPAV 180

Qy 181 LEAKATWTRDAAWQPLRRYVEDTLVVDPVLEFIAQNALDGLLYPLVYDRFVDERIAL 240
Db 181 LEAGKLAWEEDPRQGLRRYVEDSVVDDPVLEFVAQNALDGLLYPLIYGHFVDDHLAL 240

Qy 241 EGSVAAMLTAFPPEWHTESNRWIDAVVKTMAESDDNRALLARWTRDWSARAALAPV 296
Db 241 OGGTAVAILTSFPEWHDETARWIDAVIKTMANAEPEAAGNRALLSQWFSQWADRAQAA 300

Qy 297 LAPVAARALQDAGRAALDEVE-QFHARAARLGIAL 331
Db 301 LAPVAEMALGEGHAALGEVRRALDARATKLGSL 336

RESULT 4
O84959
ID O84959 PRELIMINARY; PRT; 331 AA.
AC O84959;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN POXB.
OS Ralstonia sp. E2.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RX MEDLINE; 98361023.
RA HINO S., WATANABE K., TAKAHASHI N.;
RT "Phenol hydroxylase cloned from Ralstonia eutropha strain E2 exhibits
RT novel kinetic properties.";
RL Microbiology 144:0-0(0).
DR EMBL; AF026065; AAC32453.1; -.
SQ SEQUENCE 331 AA; 37592 MW; F273FAA0 CRC32;

Query Match 48.3%; Score 832; DB 2; Length 331;
Best Local Similarity 50.0%; Pred. No. 5e-61;
Matches 166; Conservative 49; Mismatches 113; Indels 4; Gaps 3;

Qy 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOENFHYRPTWDPDYEIFD 60
Db 1 MQVDIKTQOIQPLRQTYGHVARRF-GDKPASRYQEATYDVQSEVNFHYRPTWAPQFELYD 59

Qy 61 PSRAIRMANWYALKDPQFYASWATTRARQODAMESNFESVRRMIGLMDDDVAARA 120
```

```

Db 60 KRRTAEMADWYALKDPQYYGAYVCTGRGQEEAKNEFAFVKEKRLGLQALPLEWERL 119
QY 121 LDVLPVLRHAAGANNNAOICALGCTGTPTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
Db 120 TDGLLLPLRHVEAANNMNFYCADYGMGTATQACTYCAMDLGTAQYLSRIGLLDNTG 179
QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVVELFIAQNLALDGLLYPLVYDRFVDERI 238
Db 180 AALERARAAWLESEAWQPLRRYVEHSEFVIEDWTFQTFQNLALDGLLYPLVY--QHADA 238
QY 239 ALGEGSAVAMLTAFMPPEWHTESNRWIDAVVKTMAAESDDNRALLARWTRDWSARA 298
Db 239 VRACGTGLAVVTEFMDNRDRHVRWVDVAVVQTAASEANRALLSWAGEARQAQAEALR 298
QY 299 PVAARALODAGRAALDEVROQFIHARLGI 330
Db 299 PVAAILLGGEGEQAIALCLEQFARLAKLGA 330

```

```

RESULT 5
Q52171 PRELIMINARY; PRT; 331 AA.
AC Q52171;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOOXYGENASE P1 COMPONENT).
GN PHLB.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P35X / NCBI 9869;
RX MEDLINE; 95129877.
RA NG L.C., SHINGLER V., SZE C.C., POH C.L.;
RT "Cloning and sequences of the first eight genes of the chromosomally
RT encoded (methyl) phenol degradation pathway from Pseudomonas putida
RT P35X."
RL Gene 151:29-36(1994).
CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -1- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC -1- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X79063; CAA55661.1; -.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron.
SQ SEQUENCE 331 AA; 38178 MW; FD7E5BA1 CRC32;

```

```

Query Match 45.2%; Score 778.5; DB 2; Length 331;
Best Local Similarity 46.8%; Pred. No. 1.3e-56;
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;

QY 1 MTELTVDIKPLRHFAHVAQNIIGDKTATRYQEGMGAQPOENFYHRTWDPDYEIFD 60
Db 1 MSVEIKNTVDPIRQTYGNLQRRF-GDKPASRYQEA5YDIEAVTNFYHRLWDPQHELHD 59
QY 61 PSRAIRMANWYALKDPQRYFYASWATTRARQODAMESNFEEVSRMIGLMDRDVAARA 120
Db 60 PTTAIRMTHVKVTDPRQYFYAYVQTRARMQEAETHAYCFCEKRELLSRPAELQAKL 119
QY 121 LDVLPVLRHAAGANNNAOICALGCTGTPTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
Db 120 LRLVPLRHAELGANMNNSSIAGDSIAATVTQMHIYQAMDRLMGQYLSRIGLLDGGTG 179

```

```

QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVVELFIAQNLALDGLLYPLVYDRFVDERI 238
Db 180 EALDOAKAYWLDPIWQGLRRYVEDSVIRDFELGLAQNLVLDGLLOPLMYQRF--DOWL 238
QY 239 ALGEGSAVAMLTAFMPPEWHTESNRWIDAVVKTMAAESDDNRALLARWTRDWSARA 298
Db 239 TENGSGDVAMLTETFRMDRWYGESTRWVDAFKTVLAENDANREQVQAMLEWPEPRAYEALL 298
QY 299 PVAARALODAGRAALDEVROQFIHARLGI 329
Db 299 PLABEA---TGIAALDEVRSFAFATRLQKIGL 326

```

```

RESULT 6
Q52162 PRELIMINARY; PRT; 342 AA.
AC Q52162;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOOXYGENASE P1 COMPONENT).
GN PHLB.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE; 95272534.
RA HERRMANN H., MUELLER C., SCHMIDT I., MAHNKE J., PETRUSCHKA L.,
RA HAHNKE K.;
RT "Localization and organization of phenol degradation genes of
RT Pseudomonas putida strain H."
RL Mol. Gen. Genet. 247:240-246(1995).
CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -1- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC -1- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X80765; CAA36741.1; -.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 342 AA; 39370 MW; 430445B9 CRC32;

```

```

Query Match 44.7%; Score 770.5; DB 2; Length 342;
Best Local Similarity 46.5%; Pred. No. 6.1e-56;
Matches 154; Conservative 54; Mismatches 116; Indels 7; Gaps 4;

QY 1 MTELTVDIKPLRHFAHVAQNIIGDKTATRYQEGMGAQPOENFYHRTWDPDYEIFD 60
Db 12 MSVEIKNTVDPIRQTYGNLQRRF-GDKPASRYQEA5YDIEAVTNFYHRLWDPQHELHD 70
QY 61 PSRAIRMANWYALKDPQRYFYASWATTRARQODAMESNFEEVSRMIGLMDRDVAARA 120
Db 71 PTTAIRMTHVKVTDPRQYFYAYVQTRARMQEAETHAYCFCEKRELLSRPAELQAKL 130
QY 121 LDVLPVLRHAAGANNNAOICALGCTGTPTAPAMFHAMDNGLGVAQYLTRLALAM--AEP 178
Db 131 LRLVPLRHAELGANMNNSSIAGDSIAATVTQMHIYQAMDRLMGQYLSRIGLLDGGTG 190
QY 179 DVLEAAKATWTRDAAWQPLRRYVEDTLVADPVVELFIAQNLALDGLLYPLVYDRFVDERI 238
Db 191 EALDOAKAYWLDPIWQGLRRYVEDSVIRDFELGLAQNLVLDGLLOPLMYQRF--DOWL 249
QY 239 ALGEGSAVAMLTAFMPPEWHTESNRWIDAVVKTMAAESDDNRALLARWTRDWSARA 298

```

Db 250 TENGSDVAMLTFRMDMYGSTRWVDMFKTVLAENDANREQVQAWLEWPEPRAYEALL 309  
QY 299 PVAARALQDAGRAALDEVREQFHARAALGI 329  
Db 310 PLAEER---TGIALDEVRSFAFATRLQKIGL 337  
RESULT 7  
Q43979 PRELIMINARY; PRT; 333 AA.  
AC Q43979;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE PHENOLHYDROXYLASE COMPONENT.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
OC Moraxellaceae; Acinetobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCIB8250;  
RX MEDLINE; 96154937.  
RA EHRT S., SCHIRMER F., HILLEN W.;  
RT "Genetic organization, nucleotide sequence and regulation of  
RT expression of genes encoding phenol hydroxylase and catechol 1,2-  
RT dioxygenase in Acinetobacter calcoaceticus NCIB8250.";  
RL Mol. Microbiol. 18:13-20(1995).  
DR EMBL; 236909; CA85381.1; -;  
SQ SEQUENCE 333 AA; 38382 MW; 83055EC9 CRC32;  
Query Match 37.2%; Score 640.5; DB 2; Length 333;  
Best Local Similarity 39.4%; Pred. No. 3e-45;  
Matches 127; Conservative 70; Mismatches 120; Indels 5; Gaps 4;  
QY 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYQEGMGAQPOENFHYRPTWDPDYEIFD 60  
Db 1 MTLEIKTSNLEPIQOTYAIERRFGA-KPATRYQEVSDIOASTNFHYRPLMKPKDKTLND 59  
QY 61 PPSAIRMANVYALKDPQFYFASWATTRARQODAMESNFEVESRRMIGLMRDVVAARA 120  
Db 60 KHTALQMDWYAFKDPQFYGYVQHRARLQDTAESHYAFFEKRQLVNLSDEVKOKI 119  
QY 121 LDVLPVLRHAAGANNNAQICALCYGVFTAPAMFAMDMNLGVAQYIYTRALALAM--AEP 178  
Db 120 IOCLLPFRYVEQTANLHMMSGAYGYGVITQACIFAAMDRLGMAQYISRGILIDGNTG 179  
QY 179 DVLEAAKATWTDRAAWQPLRYVEDTLVADPVELFIAQNLALDGLLYPLVYDRFVDRI 238  
Db 180 ESLOQAQKAWLNDTQWPLKCEQSLTEQDWFKLYILQNLIDSLMLQELVFGQ-LDEWL 238  
QY 239 ALEGSSAVAMLTAFMPPEWHTESNRWIDAVVKTMAESDDNRALLARWTRDSARAALAA 298  
Db 239 VENGGRDIAITEFMKOCCLTLAKWSDSVLTAISESDNKTLIQSITELLPOVKQAFS 298  
QY 299 PVAARALQDAG-RAALDEVREQ 319  
Db 299 AWAQALTDSGIDSLNKAISER 320  
RESULT 8  
O32429 PRELIMINARY; PRT; 333 AA.  
AC O32429;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE DNS OXYGENASE COMPONENT.  
GN DSOB.  
OS Acinetobacter sp.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
OC Moraxellaceae; Acinetobacter.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-20B;  
RX MEDLINE; 98005684.  
RA HORINOUCHI M., KASUGA K., NOJIRI H., YAMANE H., OMORI T.;  
RT "Cloning and characterization of genes encoding an enzyme which  
RT oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B.";  
RL FEMS Microbiol. Lett. 155:99-105(1997).  
DR EMBL; D85083; BAA23331.1; -;  
SQ SEQUENCE 333 AA; 38393 MW; BCL06309 CRC32;  
Query Match 37.2%; Score 640.5; DB 2; Length 333;  
Best Local Similarity 38.7%; Pred. No. 3e-45;  
Matches 129; Conservative 69; Mismatches 128; Indels 7; Gaps 4;  
QY 1 MTIELKTVDIKPLRHTFAHVAQNIIGDKTATRYQEGMGAQPOENFHYRPTWDPDYEIFD 60  
Db 1 MTLEIKTSNLEPIQOTYAIERRFGA-KPATRYQEVSDIOASTNFHYRPLMKPKDKTLND 59  
QY 61 PPSAIRMANVYALKDPQFYFASWATTRARQODAMESNFEVESRRMIGLMRDVVAARA 120  
Db 60 KHTALQMDWYAFKDPQFYGYVQHRARLQDTAESHYAFFEKRQLVNLSDEVKOKI 119  
QY 121 LDVLPVLRHAAGANNNAQICALCYGVFTAPAMFAMDMNLGVAQYIYTRALALAM--AEP 178  
Db 120 IOCLLPFRYVEQTANLHMMSGAYGYGVITQACIFAAMDRLGMAQYISRGILIDGNTG 179  
QY 179 DVLEAAKATWTDRAAWQPLRYVEDTLVADPVELFIAQNLALDGLLYPLVYDRFVDRI 238  
Db 180 ESLOQAQKAWLNDTQWPLKCEQSLTEQDWFKLYILQNLIDSLMLQELVFGQ-LDEWL 238  
QY 239 ALEGSSAVAMLTAFMPPEWHTESNRWIDAVVKTMAESDDNRALLARWTRDSARAALAA 298  
Db 239 VENGGRDIAITEFMKOCCLTLAKWSDSVLTAISESDNKTLIQSITELLPOVKQAFS 298  
QY 299 PVAARALQDAGRAALDEVREQFHARAALGI 331  
Db 299 AWAQALTDSG--IDSGLNKAISERSKAGTIL 328  
RESULT 9  
O87802 PRELIMINARY; PRT; 330 AA.  
AC O87802;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE TOLUENE, O-XYLENE MONOOXYGENASE OXYGENASE SUBUNIT.  
GN TOUE.  
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
OC Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OX1;  
RX MEDLINE; 98432776.  
RA BERTONI G., MARTINO M., GALLI E., BARBIERI P.;  
RT "Analysis of the gene cluster encoding toluene/o-xylene monooxygenase  
RT from Pseudomonas stutzeri OX1.";  
RL Appl. Environ. Microbiol. 64:3626-3632(1998).  
DR EMBL; AJ005663; CAA06658.1; -;  
KW Monooxygenase.  
SQ SEQUENCE 330 AA; 38334 MW; 4EFFIAD8 CRC32;  
Query Match 12.9%; Score 222.5; DB 2; Length 330;  
Best Local Similarity 25.1%; Pred. No. 8.2e-11;  
Matches 83; Conservative 52; Mismatches 147; Indels 49; Gaps 14;  
QY 10 IKPLRHRTFAHVAQNIIGDKTATRYQEGMGAQPOENFHY-----RP-----TWD 53  
Db 9 LKPLK-TWSHLA---GNRRRSEYE-----VSTNLHYFTDNPERPWELSDNLPMQWTY 58

QY 54 PDYEFPSRAIRMANWYALKDPQFYASWATTRARQDAMESFEFESRRMIGLMR 113  
 Db 59 KKY-CFD---SPLKDDNNAFRDPQLVYRTYNLQDQESYVQGLFDQLNDRGHQDMLT 114  
 QY 114 DDVAARALDVLVPLRHAAGANNMNAQICALGYCTVTAPAMFHAMDNGLGVAQYLRLA- 172  
 Db 115 REWVETTLAREYTPARYLFHALQMGSVYIHQIPASTITNCATYETAHDL---RWLTHWAY 171  
 QY 173 ----LMAAEPDV--LEAAKATWTDRDAWQPLRRYVEDTLVADPVELFIAONLALDGLLY 226  
 Db 172 RRELANCYFDVGGKRRDWEWNPDAWQGFRELIEKALTAWDGEAFTALNL-----VTK 227  
 QY 227 PLVYDFVDERIAL---EGGSAVAMLTAFMPWHITESNRWIDAYVVKMAESDDNRALLA 283  
 Db 228 PAVEEALLQGLSLAQSEGDTLLGLLAQAQKRAERHRRWSSALVK-MALEKEGNREVIQ 286  
 QY 284 RWRDWSARAEEALAPVAARALQDAGRAALD 314  
 Db 287 KWAKWEPLADKAIEAYCS-ALPDGENAIVE 316

## RESULT 10

Q51943

ID Q51943 PRELIMINARY; PRT; 329 AA.

AC Q51943;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)

DE TOLUENE-3-MONOOXYGENASE OXYGENASE SUBUNIT.

GN TBUA2.

OS Burkholderia pickettii (Pseudomonas pickettii).

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Ralstonia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PRO1.

RX MEDLINE; 95172404.

RA BYRNE A.M., KUKOR J.J., OLSEN R.H.;

RT "Sequence analysis of the gene cluster encoding toluene-3-

RT monooxygenase from Pseudomonas pickettii PKO1.";

RL Gene 154:65-70(1995).

DR EMBL; U04052; AAB09622.1; -.

KW Monooxygenase.

SQ SEQUENCE 329 AA; 37525 MW; C61B3727 CRC32;

## Query Match

Best Local Similarity 11.2%; Score 193; DB 2; Length 329;

Matches 88; Conservative 55; Mismatches 132; Indels 82; Gaps 18;

QY 10 IKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOE-----NFHYRP-----TWDPDY 56  
 Db 8 LKPLK-TWSHLA-----ARR-----RKPEYEIVSTNLHYTTNDPDPPELDPNF 51  
 QY 57 EI----FDPSSA--IRMANWYALKDPQFYASWATTRARQDAMESFEFESRRMIGL 111  
 Db 52 EMAQWFKRNRNASPLTHPDNNAFRDPDELVYRTYNML-----QDQET-----YFGL 99  
 QY 112 MRDDVAARALDVLV-----PLRHAAGANNMNAQICALGYCTVTAPAMFHA 158  
 Db 100 L-DOFSERGHDAMLERTWAGTFLARLYTPVRYLFHTLQMGSAYLTLQAPASTISNCAAYQT 158  
 QY 159 MDNLGVAQYLTRLALAAEPDVLEA-----AKATWTRDAWQPLRRYVEDTLVADPV 211  
 Db 159 ADSL---RWLTHTAIRTKELSTQTSFSDLGFGTDERRRYEQDPAMQGWKRLVEHALVANDWA 215  
 QY 212 ELFTAONLALDGLLYPLVYD---RFVDERIALEGSAVAMLTAFMPWHITESNRWIDAVV 268  
 Db 216 ECFVALSL-----VVRPAVEAVLRSLGEAARHNGDTLLGLLTDQAQADQHRHWAGALV 271  
 QY 269 KTMMAESDDNRALLARWTRWSARAEEALAPVAARALQDAGRA---ALDEVREQFHA 322  
 Db 272 R-MALEQPGNREVITGWLKWEPLADEAIVAYCS-ALPEAPAAQARATAAVREFRHS 326

## RESULT 11

O69182

ID O69182 PRELIMINARY; PRT; 332 AA.

AC O69182;

DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)

DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)

DE PUTATIVE HYDROXYLASE COMPONENT.

GN PHLO.

OS Alcaligenes eutrophus.

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Ralstonia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JMP134;

RA HARKER A.R., AYOUBI P.J.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF065891; AAC77384.1; -.

SQ SEQUENCE 332 AA; 37511 MW; 871144CB CRC32;

## Query Match

Best Local Similarity 11.1%; Score 190.5; DB 2; Length 332;

Matches 82; Conservative 53; Mismatches 145; Indels 47; Gaps 13;

QY 10 IKPLRHTFAHVAQNIIGDKTATRYOEGMGAQPOE-----NFHYRP-----TWDPDY 56  
 Db 8 LKPLK-TWSHLA-----ARR-----RKPEYEIVSTNLHYTTNDPDPPELDPNF 51  
 QY 57 EI----FDPSSA--IRMANWYALKDPQFYASWATTRARQDAMESFEFESRRMIGL 111  
 Db 52 EMAQWFKRNRNASPLKHAADNNAFRDPDELVYRTYNMLQDQETVSGLLDQFSERGHDSM 111  
 QY 112 MRDDVAARALDVLVPLRHAAGANNMNAQICALGYCTVTAPAMFHAMDNGLGVAQYLRL 171  
 Db 112 LEHSWAGTFLARLYTPARYLFHALQMGSAYLTLQMAPASTISNCAAYQTADSL---RWLTHT 168  
 QY 172 A-----LMAAEPDVLEA--KATWTRDAWQPLRRYVEDTLVADPVELFIAONLALDGL 224  
 Db 169 AYRTRELSTQTFDVGFGTDERRRYEQDPAMQGWKRLVEHALVANDWAECFVAFSLVRPA 228  
 QY 225 LYPLVYDFVDERIALEGSAVAMLTAFMPWHITESNRWIDAVVKTMAESDDNRALLAR 284  
 Db 229 MEEAVL-RGEGEAARHNGDTLLGLLTDQAQADQHRHWAGALVR-MALETTPGNRDLVAG 286  
 QY 285 WTRDWSARAEEALAPVAARALQDAGRA 311  
 Db 287 SIARWAPLADDAISAYCA-ALPDAPNA 312

## RESULT 12

Q53025

ID Q53025 PRELIMINARY; PRT; 343 AA.

AC Q53025;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)

DE EPOXIDASE SUBUNIT.

GN AMOA.

OS Nocardia corallina.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-B-276;

RA SAEKI H., FURUHASHI K.;

RT "Cloning and characterization of a Nocardia corallina B-276 gene

RT cluster encoding alkene monooxygenase.";

RL J. Ferment. Bioeng. 78:339-406(1994).

DR EMBL; D37875; BAA07112.1; -.

SQ SEQUENCE 343 AA; 38462 MW; 9195C333 CRC32;

Query Match	9.7%	Score 167.5;	DB 2;	Length 341;
Best Local Similarity	26.1%	Pred. No. 2.9e-06;		
Matches 74;	Conservative 49;	Mismatches 124;	Indels 37;	Gaps 14;



RX MEDLINE; 97288080.  
RA MCDONALD I.R., UCHIYAMA H., KAMBE S., YAGI O., MURRELL J.C.;  
RT "The soluble methane monooxygenase gene cluster of the  
RL trichloroethylene-degrading methanotroph Methylocystis sp. strain M.";  
DR Appl. Environ. Microbiol. 63:1898-1904(1997).  
KW EMBL; U81594; AAC45290.1; -;  
SQ MONOOXYGENASE.  
KW SEQUENCE 395 AA; 45170 MW; E4507FBD CRC32;

Query Match 7.1%; Score 121.5; DB 2; Length 395;  
Best Local Similarity 20.8%; Pred. No. 0.022;  
Matches 68; Conservative 50; Mismatches 164; Indels 45; Gaps 12;  
QY 28 KTATRYQEGMGAQPO-----ENFH-YRPTWDPDYEIFDPSPSAIRMANWY 72  
DB 47 KRUSEYQSLSCYAQPNDWIAGGLDWDGWTOKFHGGRPSWGNE-----STELRTTDWY 99  
QY 73 ALKDPQFYASWA---TTRARQDAMESNPEFVESRRMIG-LMRDDVAARALDVLVPLR 128  
DB 100 RHRDPARRWHAPYVKDSEEARYTQRLAAYSSEGSIRTIDPYWRDEILNKYYGALLYNE 159  
QY 129 HAANGANNNAQICALGYGTFTAPAMFAMDNLGVAQ--YLTRLALAMAEPDV---LEA 183  
DB 160 YGLENAHSSVGRDC---LSDTIROSATFAGLDKVDNAQMIQMERLFIKLVPGFDASTDV 216  
QY 184 AKATWTRDAWQPLRRYVEDTLV-VADPVELFIAQNALDGLLYPLVYDRFVDERIALEG 242  
DB 217 PKIWTSDPIYAGARGAVEEIQGIQDWNELIHWAGHAVYDATFGQFARREFQRLATVYG 276  
QY 243 GSAVAMLTAFMPEWHTESNRWIDAVVKTMAESDD----NRALLARWTRDWSARAAL- 297  
DB 277 DTLTPFTAQSOITYFTQTRGAIDDLFYVCLANDPEFGAHHNRTFLNATHEYLARSVTALK 336  
QY 298 --APVAARALQDAG---RAALDEVREQ 319  
DB 337 DFVGIIYAKVEKVAGATDRAGVSEALQR 363

Search completed: September 26, 2000, 20:28:52  
Job time: 1011 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:37 ; Search time 60.65 seconds  
(without alignments)  
22.495 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451

Sequence: 1 MSNFIARQANEDSRPIVDA.....ITLSGHIDEEDDEFLSMWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*

4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*

5: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	355	78.7	92	1	US-08-319-387-3
2	122	27.1	103	5	5171684-5
3	65	14.4	1162	2	US-08-663-566A-15
4	65	14.4	1162	2	US-08-023-610-15
5	65	14.4	1162	2	US-08-288-065A-15
6	65	14.4	1162	2	US-08-362-240A-15
7	65	14.4	1162	4	PCT-US95-10245-15
8	64	14.2	292	3	PCT-US85-1190-3
9	62.5	13.9	406	5	5212296-6
10	62	13.7	322	2	US-08-505-218-6
11	62	13.7	897	1	US-07-960-389-2
12	61	13.5	256	3	US-08-306-769-89
13	61	13.5	256	4	PCT-US95-14442A-89
14	59.5	13.2	339	1	US-08-266-451B-22
15	59.5	13.2	339	2	US-08-748-725-22
16	59.5	13.2	464	1	US-08-475-894-4
17	59.5	13.2	464	2	US-08-484-710-4
18	59.5	13.2	464	2	US-08-484-709-4
19	59.5	13.2	553	1	US-08-475-894-2
20	59.5	13.2	553	1	US-08-484-710-2
21	59.5	13.2	553	2	US-08-484-709-2
22	58	12.9	117	1	US-08-499-215-3
23	58	12.9	682	2	US-08-436-900A-4
24	58	12.9	1477	3	US-08-492-459-10
25	58	12.9	1477	3	US-08-423-752-10
26	58	12.9	1477	3	US-08-945-994-3
27	55.5	12.3	381	2	US-08-387-942C-25
28	55.5	12.3	425	1	US-08-190-802A-48

29 55.5 12.3 425 3 US-08-105-454-10 Sequence 10, Appli  
30 55.5 12.3 575 2 US-08-766-858A-5 Sequence 5, Appli  
31 55.5 12.3 1403 2 US-08-387-942C-3 Sequence 3, Appli  
32 55.5 12.3 2544 2 US-08-576-626A-32 Sequence 32, Appli  
33 55 12.2 308 2 US-08-484-905-78 Sequence 78, Appli  
34 55 12.2 308 3 US-08-481-985B-78 Sequence 78, Appli  
35 55 12.2 484 1 US-08-597-236-2 Sequence 2, Appli  
36 55 12.2 484 1 US-08-746-682A-2 Sequence 2, Appli  
37 55 12.2 672 1 US-08-049-254-2 Sequence 2, Appli  
38 55 12.2 672 1 US-08-472-934-2 Sequence 2, Appli  
39 55 12.2 672 2 US-08-323-460A-2 Sequence 2, Appli  
40 55 12.2 672 2 US-08-461-145C-2 Sequence 2, Appli  
41 55 12.2 672 3 US-08-461-145C-2 Sequence 2, Appli  
42 54 12.0 115 2 US-08-487-031-15 Sequence 15, Appli  
43 54 12.0 853 3 US-08-451-946B-6 Sequence 6, Appli  
44 54 12.0 853 3 US-08-446-938B-6 Sequence 6, Appli  
45 54 12.0 853 3 US-08-446-939B-6 Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-319-387-3  
; Sequence 3, Application US/08319387  
; Patent No. 5543317  
; GENERAL INFORMATION:  
; APPLICANT: Shields, Malcolm S.  
; APPLICANT: Francesconi, Stephen C.  
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,  
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/319,387  
; APPLICATION NUMBER: US/08/319,387  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/167,457  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694,718  
; FILING DATE: 02-MAY-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UWF-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 92 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-319-387-3





INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-10245-15

Query Match 14.4%; Score 65; DB 4; Length 1162;  
Best Local Similarity 26.7%; Pred. No. 6.7;  
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VFTAFQANEDSRPIVDAIVA-----DNRAVVVSPGMVKIDAPDRLTIR 48  
Db 934 VFTHSYTPDSFVNVTAVGFCVKPANASQYAIVPANGRGIFQVNGSYIITARDMYMR 993  
QY 49 RETIEL-TGTRPDLOQLQNLITLSGHIDDD-----DEFTLSWS 88  
Db 994 AITAGDIVTLTSCQANYSVNKTIVITTFVNDNDDFENDELSKWN 1038

RESULT 8  
US-08-851-190-3  
Sequence 3, Application US/08851190  
Patent No. 6074843

GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: HUMAN TSC-22-LIKE PROTEIN  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,190  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0285 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 292 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TONSNOT01  
CLONE: 736663  
US-08-851-190-3

Query Match 14.2%; Score 64; DB 3; Length 292;  
Best Local Similarity 24.6%; Pred. No. 1.3;  
Matches 17; Conservative 13; Mismatches 21; Indels 18; Gaps 2;

QY 13 DSRPIVDAVNPRRAVVVSPGMVKIDAPDRLTIRRETIEELTGTRFDLQQLVNLITL 72  
Db 20 DSRPSPFALYFTTHDASLVHKSP-----DPFGAVAAQKFSLAH---SMLAI 61  
QY 73 SGHIDEDDD 81  
Db 62 SGHLSDSD 70

RESULT 9  
5212296-6  
Patent No. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.  
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO: 6:  
LENGTH: 406  
5212296-6

Query Match 13.9%; Score 62.5; DB 5; Length 406;  
Best Local Similarity 35.5%; Pred. No. 3.2;  
Matches 22; Conservative 9; Mismatches 28; Indels 3; Gaps 3;  
QY 19 DAIVADNPR-AVVVESP-GMVKIDAPDRLTIRRETIEELTGTRFDLQQLVNLITLSGHI 76  
Db 76 DNFPAISPFREAVRESPQAFICGLDPHCRTTRMTISEFTVRIKGMREVEEV-VHGFL 134  
QY 77 DE 78  
Db 135 DE 136

RESULT 10  
US-08-505-218-6  
Sequence 6, Application US/08505218  
Patent No. 5914447  
GENERAL INFORMATION:  
APPLICANT: ARAYA, ALEJANDRO  
APPLICANT: MOURAS, ARMAND  
TITLE OF INVENTION: TRANSGENIC PLANTS INCLUDING A HYBRID  
NUCLEIC ACID SEQUENCE, COMPRISING AN UNEDITED  
MITOCHONDRIAL GENE FRAGMENT FROM HIGHER PLANTS AND PROCESS  
FOR PRODUCING THEM  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/505,218  
FILING DATE: 03-NOV-1995  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: KRAUS, ERIC J  
REGISTRATION NUMBER: 36,190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 684-1111  
TELEFAX: (703) 684-1124  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 322 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-505-218-6

Query Match 13.7%; Score 62; DB 2; Length 322;  
Best Local Similarity 32.1%; Pred. No. 2.7;  
Matches 27; Conservative 12; Mismatches 25; Indels 20; Gaps 6;  
QY 13 DSRPIVDA-----IV--ADNPRAVVVSPGMVKIDA-PDR-----LTIRRETI-----E 53  
Db 229 DNRVVVFAKTHLRMIVTADVPHSWAVPSSG-VKCDVAPGRSLTFTISVOREGVYGGCS 287  
QY 54 ELTGTRFDLQLOLVNLTLSGHID 77  
Db 288 EIRCTNHAFPIVVEAVTLKDYAD 311

RESULT 11  
US-07-960-389-2  
Sequence 2, Application US/07960389  
Patent No. 5705611  
GENERAL INFORMATION:  
APPLICANT: HAYASHIDA, Kasuhiro;  
TITLE OF INVENTION: Human GM-CSF Receptor Component  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schering-Plough Corporation  
STREET: 2000 Galloping Hill Road  
CITY: Kenilworth  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07033  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disc  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: System Software 7.1  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07960,389  
FILING DATE: 07-JAN-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 554,745  
FILING DATE: 18-JUL-1990  
APPLICATION NUMBER: PCT/US 91/04846  
FILING DATE: 16-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Blasdale, John H. C.  
REGISTRATION NUMBER: 31,895  
REFERENCE/DOCKET NUMBER: DX0143Q  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 298-2902  
TELEFAX: (908) 298-5388  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 897 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:

OTHER INFORMATION: Human GM-CSF receptor; Signal Sequence: -17 to -1; Tran  
US-07-960-389-2  
Query Match 13.7%; Score 62; DB 1; Length 897;  
Best Local Similarity 24.6%; Pred. No. 11;  
Matches 28; Conservative 9; Mismatches 31; Indels 46; Gaps 5;  
QY 3 NVFTAFQANEDSRPIYDAIVADN-----PRAVVVSPGMVKIDA-----PDR 44  
Db 58 NVTILRRVNEDLLEFVSCDLSDDMPWSACPHRCVPRCVFVVTVDYFSFQDR 117  
QY 45 LTIRRETIETGTGRFDLQLOLVNLTLSGH-----IDEDDDFTLSWS 88  
Db 118 -----PLGTRL-----TVILTQHVQPPEDRLQISTDDQDFILTWS 153

RESULT 12  
US-08-906-769-89  
Sequence 89, Application US/08906769  
Patent No. 6077687  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,769  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-769-89

Query Match 13.5%; Score 61; DB 3; Length 256;  
Best Local Similarity 22.9%; Pred. No. 2.6;  
Matches 16; Conservative 16; Mismatches 26; Indels 12; Gaps 1;  
QY 25 NPRAVVVESP-----GMVKIDAPDRLTIRRETIETGTGRFDLQLOLVNLT 72



```

Db      90 HPXAIIKHYPAYGNVTDIDMEXALIKVRRPRLNRTVRTVKLTDVGKDMPSGELATVTG 149
QY      73 SGHIDEEDDE 82
       I::|||:
Db      150 WGNLGEDEDD 159

RESULT 13
PCT-US95-14442A-89
; Sequence 89, Application PC/TUS9514442A
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Arfsten, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 256 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-14442A-89

Query Match          13.5%; Score 61; DB 4; Length 256;
Best Local Similarity 22.9%; Pred. No. 2.6;
Matches 16; Conservative 16; Mismatches 26; Indels 12; Gaps 1;

QY      25 NPRAVVVESP-----GMVKIDAPDRLTIRRETTIEELTGTRFDLQQQLQVNLI 72
       :|::|::|:
Db      90 HPXAIIKHYPAYGNVTDIDMEXALIKVRRPRLNRTVRTVKLTDVGKDMPSGELATVTG 149
QY      73 SGHIDEEDDE 82
       I::|||:
Db      150 WGNLGEDEDD 159

RESULT 14
US-08-266-451B-22
; Sequence 22, Application US/08266451B

```

;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/748,725  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/266,451  
; FILING DATE: 23-June-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lech, Karen F.  
; REGISTRATION NUMBER: 35,238  
; REFERENCE/DOCKET NUMBER: 00786/219002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 339  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-748-725-22

Query Match 13.2%; Score 59.5; DB 2; Length 339;  
Best Local Similarity 25.6%; Pred. No. 6.2;  
Matches 23; Conservative 14; Mismatches 30; Indels 23; Gaps 3;  
QY 15 RPIVDAIVADNPRAVV-----ESPGMWKIDAPDRLTI-----RRETI 52  
Db 159 KPILDEIDAGGSAAMKYWNPVVKLKGAMGMPVAGLPDQTVSAEPEVAEGEEREESI 218  
QY 53 EELGTREDFLQQLQVNLITLSGHIDEDDE 82  
Db 219 VHQFASLGDEGLKALAS-GGNKDEDESE 247

Search completed: September 26, 2000, 20:22:40  
Job time: 3814 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:12 ; Search time 82.1 Seconds  
(without alignments)  
67.086 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451

Sequence: 1 MSNVIFAQANEDSRPIVDA.....ITLSGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_64:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	54.8	90	1 S44305	phenol 2-monooxyge
2	247	54.8	90	1 C37831	phenol 2-monooxyge
3	232	51.4	89	1 S47289	phenol 2-monooxyge
4	83.5	18.5	862	2 T36380	probable large ATP
5	77	17.1	1684	2 T02632	hypothetical prote
6	70.5	15.6	174	2 G69411	small heat shock p
7	69	15.3	822	2 S77112	DNA mismatch repal
8	68.5	15.2	143	2 D75617	response regulator
9	67.5	15.0	410	2 JC4287	cytochrome P450 sc
10	66.5	14.7	2277	2 B47648	perlecan homolog u
11	66.5	14.7	2295	2 T19820	hypothetical prote
12	66.5	14.7	2482	2 T19819	hypothetical prote
13	66.5	14.7	2482	2 A47648	perlecan homolog u
14	66.5	14.7	3375	2 T19821	hypothetical prote
15	66	14.6	138	2 D39049	methane monooxygen
16	66	14.6	181	2 A40607	monofunctional cho
17	66	14.6	294	2 D81430	UDP-3-O-[3-hydroxy
18	66	14.6	390	2 C69276	coenzyme F420--qui
19	66	14.6	619	2 A43361	Ets-related transc
20	66	14.6	649	2 S74823	N-acetylmutamoyl-L
21	65	14.4	1162	2 S07421	E2 glycoprotein pr
22	65	14.4	1232	2 S40766	hypothetical prote
23	64.5	14.3	316	2 B47157	magnesium transpor
24	64.5	14.3	452	2 I49595	cytokeratin 15 - m
25	64	14.2	347	2 A75537	hypothetical prote
26	64	14.2	487	1 BWSOGM	gtfA protein - str
27	64	14.2	1162	2 S14939	E2 glycoprotein pr
28	63.5	14.1	224	2 G69783	transcription regu
29	63.5	14.1	394	2 T06180	methionine adenosy

30	63.5	14.1	420	1 DCECD	diaminopimelate de
31	63	14.0	366	2 S66351	methionine adenosy
32	63	14.0	557	2 S21596	extracellular prot
33	63	14.0	828	2 S56250	probable membrane
34	62.5	13.9	139	2 A69052	riboflavin synthas
35	62.5	13.9	205	2 S39863	late competence op
36	62.5	13.9	360	2 T06592	methionine adenosy
37	62.5	13.9	406	2 A35401	cytochrome p450 10
38	62.5	13.9	964	2 T21865	hypothetical prote
39	62.5	13.9	1541	2 T30227	pipicolate-incorpo
40	62.5	13.9	4077	3 T17484	hypothetical prote
41	62	13.7	80	2 D32338	dexA.1 protein - p
42	62	13.7	166	2 I69005	histocompatibility
43	62	13.7	260	1 OBWT2	cytochrome-c oxida
44	62	13.7	260	1 S54306	cytochrome-c oxida
45	62	13.7	260	2 T01704	cytochrome-c oxida

#### ALIGNMENTS

RESULT 1

S44305

phenol 2-monooxygenase (EC 1.14.13.7) component M - Pseudomonas putida

N;Alternate names: phenolhydroxylase chain C

C;Species: Pseudomonas putida

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C;Accession: D58972; S54766; S44305; S47416

R;Ng, L.C.; Shingler, V.; Size, C.C.; Poh, C.L.

Gene 151, 29-36, 1994

A;Title: Cloning and sequences of the first eight genes of the chromosomally encoded

A;Reference number: A58972; MUID:95129877

A;Accession: D58972

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-90 <NGL>

A;Cross-references: EMBL:X79063; NID:q483477; PIDN:CAA55662.1; PID:q483480

A;Experimental source: strain P35X (NCBI 9869)

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994

R;Herrmann, H.; Mueller, C.; Schmidt, I.; Mahnke, J.; Petruschka, L.; Hahnke, K.

Mol. Gen. Genet. 247, 240-246, 1995

A;Title: Localization and organization of phenol degradation genes of Pseudomonas put

A;Reference number: S54761; MUID:95272534

A;Accession: S54766

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-41,'V',43-55,'K',57-76,'V',78-90,'N' <HE2>

A;Cross-references: EMBL:X80765; NID:g527546; PIDN:CAA56742.1; PID:g527549

A;Experimental source: strain H

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

C;Genetics:

A;Gene: phlC; phhM

C;Superfamily: phenol 2-monooxygenase component M

C;Keywords: oxidoreductase

Query Match 54.8%; Score 247; DB 1; Length 90;  
Best Local Similarity 52.3%; Pred. No. 8.2e-19;  
Matches 46; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 2 SNVFIAQANEDSRPIVDAIVADNPRAVVVSPGMVKIDAPDRLTIRRETEELTGTRFD 61

DB 3 SLVFIAQDNDAVYLAIQDNPFAVQVHHFAMIAEAKRLEIRRETVENLGRWD 62

QY 62 LQLQVNLITLSGHIDEDEFTLSWSH 89

DB 63 VQEMLDVITIGGNIDEEDDFVLEWKN 90

RESULT 2

C37831

phenol 2-monooxygenase (EC 1.14.13.7) chain P2 - Pseudomonas sp. (strain CF600)

C;Species: Pseudomonas sp.

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

A:Accession: C37831

R:Nordlund, I.; Powlowski, J.; Shingler, V.

J. Bacteriol. 172, 6826-6833, 1990

A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol

A:Reference number: A37831; MUID:91072230

A:Accession: C37831

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-90 <NOR>

A:Cross-references: GB:M60276; GB:M37764; NID:g151449; PIDN:AAA25941.1; PID:g151452

C:Superfamily: phenol 2-monooxygenase component M

C:Keywords: oxidoreductase

Query Match 54.8%; Score 247; DB 1; Length 90;

Best Local Similarity 51.1%; Pred. No. 8.2e-19;

Matches 45; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

QY 2 SNVTFATQANEDSRPDAIVADNPRAVVGSPGMVKIDAPDLRTIRRETEELTGTRFD 61

DB 3 SLVYTFQDNDAWVVEALIQDNPHAVVQHHPAMIRIEAEKREIRRETVEENLGRWD 62

QY 62 LQQLQVNLITLGGHIDDDDEFTLSWSH 89

DB 63 VQELVDVITIGGNVDEDDRFVLEKN 90

RESULT 3

S47289

phenol 2-monooxygenase (EC 1.14.13.13) chain mopM - Acinetobacter calcoaceticus

N:Alternate names: phenol hydroxylase

C:Species: Acinetobacter calcoaceticus

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S70082; S47289

R:Ehrt, S.; Schirmer, F.; Hillen, W.

Mol. Microbiol. 18, 13-20, 1995

A:Title: Genetic organization, nucleotide sequence and regulation of expression of genes

A:Reference number: S70080; MUID:96154937

A:Accession: S70082

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-89 <EH2>

A:Cross-references: EMBL:Z36909; NID:g535279; PIDN:CAA85382.1; PID:g535282

A:Experimental source: strain NCIB9250

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994

C:Genetics:

A:Gene: mopM

C:Superfamily: phenol 2-monooxygenase component M

C:Keywords: aromatic hydrocarbon catabolism; oxidoreductase

Query Match

Best Local Similarity 51.4%; Score 232; DB 1; Length 89;

Matches 40; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

QY 2 SNVTFATQANEDSRPDAIVADNPRAVVGSPGMVKIDAPDLRTIRRETEELTGTRFD 61

DB 3 SKVYALQNDTSRVIIIEQDNPEATIQYLPAMIRVESTGELVVRVETVSEKLGQWMD 62

QY 62 LQQLQVNLITLGGHIDDDDEFTLSWS 88

DB 63 IQELQNLNITLGGVDEDDDSFTLKN 89

RESULT 4

T36380

probable large ATP-binding protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T36380

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: Z21573

A:Accession: T36380

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-862 <OLI>

A:Cross-references: EMBL:AL049628; PIDN:CAB40869.1; GSPDB:GN00070; SCOEDB:SCE94.20

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCE94.20

Query Match

Best Local Similarity 18.5%; Score 83.5; DB 2; Length 862;

Matches 18; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 27 RAVVSPGMVKIDAPDLRTIRRETEELTGTRFDLQQLQVNLITLGGHIDDDDEFTLS 86

DB 556 RALLRS-GLLRPAEGRVDFLHRTFDYLGARLAVQEMDFLLVNHHLDEWDVVILLA 614

QY 87 WSH 89

DB 615 LAH 617

RESULT 5

T02632

hypothetical protein DJ1186C01.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C:Accession: T02632

R:Bemis, G.; Rohlfing, T.; Morris, M.

submitted to the EMBL Data Library, July 1998

A:Description: The sequence of Homo sapiens PAC clone DJ1186C01.

A:Reference number: Z14682

A:Accession: T02632

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1684 <BEM>

A:Cross-references: EMBL:AC004991; NID:g3342751; PIDN:AC27675.1; PID:g3342752

C:Genetics:

A:Map position: 7

A:Note: intron positions not resolved (incomplete sequence)

A:Note: WUGSC:H\_DJ1186C01.1

Query Match

Best Local Similarity 17.18%; Score 77; DB 2; Length 1684;

Matches 24; Conservative 20; Mismatches 24; Indels 22; Gaps 4;

QY 2 SNVTFATQANEDSRPDAIVADNPRAVVGSPG----MVKIDAPDLRTIRRETEELTG 57

DB 1250 SRPWFSLKHKDARPLRDVLAANPNRFTLLPFGAQTAVRPGSPSTSTM----- 1299

QY 58 TRFDLQQLQVNLITLGGHIDDDDEFTLSW 87

DB 1300 -RLDLQQAIIISI---IVKNDQ----SW 1321

RESULT 6

G69411

small heat shock protein (hsp20-1) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998

C:Accession: G69411

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343

A:Accession: G69411







## RESULT 14

T19821

hypothetical protein ZC101.2e - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T19821; T27490

R:Baynes, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19182

A:Accession: T19821

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3375 &lt;WIL&gt;

A:Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e

A:Experimental source: clone C38C6

R:Percy, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z20375

A:Accession: T27490

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3375 &lt;WIL&gt;

A:Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e

A:Experimental source: clone ZC101

C:Genetics:

A:Gene: CESP:ZC101.2e

A:Map position: 2

A:Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3

2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2

## Query Match

Best Local Similarity 14.7%; Score 66.5; DB 2; Length 3375;

Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;

QY 16

PVTDAIVADNPRAVVGSPGMVKIDAPDRLTIRTEETLTGTRFDLQ-----QLQVN 68

Db 1578 PTTNRVPESNPARYVKSPIRPIIDPAE-----QIVPE--GSFFKIRCVPGHPVSQVLT 1629

QY 69

LITLSGHIDEDDDE 82

Db 1630 FRRVSGQLNEADE 1643

## RESULT 15

D39049

methane monooxygenase (EC 1.14.13.25) component B - Methylosinus trichosporium (strain C

C:Species: Methylosinus trichosporium

C:Date: 23-Aug-1991 #sequence\_revision 26-Apr-1996 #text\_change 08-Oct-1999

C:Accession: S15209; D39049

R:Cardy, D.L.N.; Laidler, V.; Salmond, G.P.C.; Murrell, J.C.

Mol. Microbiol. 5, 335-342, 1991

A:Title: Molecular analysis of the methane monooxygenase (MMO) gene cluster of Methylosi

A:Reference number: S15207; MUID:91251762

A:Accession: S15209

A:Molecule type: DNA

A:Residues: 1-138 &lt;CAR&gt;

A:Cross-references: EMBL:X55394; NID:g44613; PIDN:CAA39070.1; PID:g44616

R:Fox, B.G.; Liu, Y.; Dege, J.E.; Lipscomb, J.D.

J. Biol. Chem. 266, 540-550, 1991

A:Title: Complex formation between the protein components of methane monooxygenase from

A:Reference number: A39049; MUID:91093180

A:Accession: D39049

A:Status: preliminary

A:Molecule type: protein

A:Residues: 2-32 &lt;FOX&gt;

C:Genetics:

A:Gene: mmoB

C:Keywords: oxidoreductase

Query Match 14.6%; Score 66; DB 2; Length 138;

Best Local Similarity 25.5%; Pred. No. 6.6;

Matches 24; Conservative 12; Mismatches 46; Indels 12; Gaps 2;

QY 2

SNVFIAFOANEDSRPIVDAIVAD-----NPRVVVSPGMVKIDAPDRLTIRRETI 52

Db 35 SNAVVLVLMKSDE---IDAIIEDIVLKGKAKNPISIVVEDKAGFWIKADGATEIDAAEA 91

QY 53

EELTGTRFDLQQLQVNLITLSGHIDEDDDEFTLS 86

Db 92 GELIGKPFVYDLLINVSSTVGRAYTLGTRETTIT 125

Search completed: September 26, 2000, 20:24:16

Job time: 3640 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein ~ protein search, using sw model

Run on: September 26, 2000, 20:22:44 ; Search time 60.65 Seconds  
(without alignments)  
29.824 Million cell updates/sec

Title: US-09-430-029-6  
Perfect score: 666  
Sequence: 1 MAVIALKPYDFPVKDAVEK.....HKDLISFRTPGLDGLGASG 118

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues  
Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	370	55.6	93	1	US-08-319-387-5
2	78.5	11.8	495	1	US-07-841-997A-2
3	78.5	11.8	495	1	US-08-290-301-2
4	78.5	11.8	495	2	US-08-588-983-2
5	78.5	11.8	495	2	US-08-588-976-2
6	71.5	10.7	273	1	US-08-221-750A-9
7	62	9.3	655	2	US-08-272-255-18
8	62	9.3	655	4	PCT-US95-08565-18
9	61	9.2	336	2	US-08-997-080-156
10	61	9.2	336	2	US-08-997-362-156
11	61	9.2	495	2	US-09-132-619-2
12	61	9.2	495	3	US-09-282-803B-2
13	61	9.2	541	2	US-09-132-619-6
14	61	9.2	541	3	US-09-282-803B-6
15	59.5	8.9	495	2	US-08-749-902-3
16	59.5	8.9	500	2	US-09-132-619-8
17	59.5	8.9	500	3	US-09-282-803B-8
18	59.5	8.9	681	2	US-08-272-255-6
19	59.5	8.9	681	2	US-08-272-255-9
20	59.5	8.9	681	4	PCT-US95-08565-6
21	59.5	8.9	681	4	PCT-US95-08565-9
22	59.5	8.9	726	2	US-08-313-185-49
23	59.5	8.9	726	2	US-08-459-499-13
24	59	8.9	267	1	US-07-928-611-13
25	59	8.9	267	2	US-08-487-811A-13
26	59	8.9	267	4	PCT-US93-07370-13
27	59	8.9	467	1	US-08-056-051-6
28	59	8.9	467	1	US-07-928-611-22

29	59	8.9	467	2	US-08-487-811A-22
30	59	8.9	467	4	PCT-US93-07370-22
31	59	8.9	542	1	US-07-814-964-13
32	59	8.9	542	1	US-08-258-442-13
33	59	8.9	542	1	US-08-328-809-8
34	59	8.9	542	4	PCT-US92-11107-13
35	59	8.9	558	4	PCT-US93-03027-2
36	58.5	8.8	360	1	US-08-205-506A-2
37	58.5	8.8	360	4	PCT-US94-02389-2
38	58	8.7	482	1	US-07-792-885A-1
39	58	8.7	482	1	US-08-142-439A-7
40	58	8.7	482	2	US-08-869-477-7
41	58	8.7	549	1	US-08-494-168-8
42	58	8.7	1298	2	US-08-690-473-2
43	58	8.7	1319	2	US-08-290-731C-2
44	58	8.7	1336	2	US-08-290-731C-6
45	57.5	8.6	902	1	US-08-396-479B-6

ALIGNMENTS

RESULT 1  
US-08-319-387-5  
; Sequence 5, Application US/08319387  
; Patent No. 554317  
; GENERAL INFORMATION:  
; APPLICANT: Shields, Malcolm S.  
; APPLICANT: Francesconi, Stephen C.  
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,  
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/319,387  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/167,457  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694,718  
; FILING DATE: 02-MAY-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UWF-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-319-387-5

Query match	11.8%	Score	76.37	Ds	17	Length	4937
Best Local Similarity	24.7%	Pred. No.	0.34				
Matches	23	Conservative	19	Mismatches	28	Indels	23
Gaps							4

  

QY	14	KDAVEKFPAPLLYVCVENHLMFPAPCLPLPDPMPFGALAGDVLPPVGYGHPDFAKIDWD	73
		:       : :	
		:       : :	
Db	80	KOLLEKFNAPVIFLSD-----IADLHYNGFSNLSILPLFHYHP--GEINFD	124
		:       : :	
		:       : :	
QY	74	RVWEF---RSGEPPWAPDPAKSLAGNGLGHKDLI	103
		:       : :	
		:       : :	
Db	125	ENAWFGYNEANQTETNEIAKT-----MHNNDLI	152



Query Match	9.38;	Score 62;	DB 2;	Length 655;
Best Local Similarity	39.68;	Pred. No. 39;		

RESULT 9  
US-08-937-080-156  
; Sequence 156, Application US/08997080  
; Patent No. 5968524  
; GENERAL INFORMATION:  
; APPLICANT: WATSON, JAMES D.  
; APPLICANT: TAN, PAUL L.J.  
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Sockman

RESULT 10  
US-08-997-362-156  
; Sequence 156, Application US/08997362  
; Patent No. 5985287  
; GENERAL INFORMATION:  
; APPLICANT: Tan, Paul  
; APPLICANT: Hiyama, Jun  
; APPLICANT: Visser, Elizabeth  
; APPLICANT: Skinner, Margot  
; APPLICANT: Scott, Linda  
; APPLICANT: Prestidge, Ross  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:

Query Match 9.2%; Score 61; DB 2; Length 495;  
Best Local Similarity 22.3%; Pred. NO. 36;  
Matches 27; Conservative 17; Mismatches 37; Indels 40; Gaps 6;





```
QY      4 IALKPVDFF-VKDAVEKFAPLLYVCWENHLMFPAPCLPLPDMFPGALAGDVLPPVYG 62
          :||| | :|:||| - - - - - |||
Db     289 LALPPYLTPDARDLVKKF-----LKNRFSQRIGGGPGDAAD--VQ 326
```

```

03 IHPDFAKJLDVWVFKEPSEWAP 86
|||::||:|:|:|
Db 327 RHPEFRHMNDLAWRVDPFPR 350

Search completed: September 26, 2000, 20:22:46
Job time: 3820 sec
```

Query Match 8.9%; Score 59.5; DB 2; Length 495;  
Best Local Similarity 26.2%; Pred. No. 53;  
Matches 22; Conservative 8; Mismatches 31; Indels 2



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:21 ; Search time 82.1 seconds  
(without alignments)  
88.945 Million cell updates/sec

Title: US-09-430-029-6  
Perfect score: 666  
Sequence: 1 MAVIALKPYDFPVKDAVEKF.....HKDLISFRTPLDGLGGASF 118

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64: \*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215.5	32.4	119	1 E37831	phenol 2-monooxyge
2	212.5	31.9	119	1 S44307	phenol 2-monooxyge
3	143	21.5	120	1 S47291	phenol 2-monooxyge
4	127	19.1	671	2 A45730	phenol 2-monooxyge
5	84.5	12.7	271	2 S46295	chlorophyll a/b-bi
6	78.5	11.8	487	1 S52261	NADH dehydrogenase
7	72.5	10.9	495	2 S34979	alpha,alpha-trehal
8	71.5	10.7	273	2 D47301	VirB9 homolog - Bo
9	71.5	10.7	640	2 S69546	phosphoenolpyruvat
10	71	10.7	522	2 S52216	viral proteinase -
11	68	10.2	201	2 S26045	sex-determining pr
12	68	10.2	482	2 T40809	conserved hypothet
13	67.5	10.1	372	2 S60207	fomP protein - Str
14	66.5	10.0	94	2 S61459	heat shock transcr
15	66.5	10.0	102	2 G25035	hypothetical prote
16	66.5	10.0	345	2 S12788	hypothetical prote
17	66.5	10.0	476	2 T43464	transcription fact
18	66.5	10.0	691	2 T46476	hypothetical prote
19	66	9.9	405	2 T16836	hypothetical prote
20	66	9.9	448	2 T06076	proline-rich prote
21	66	9.9	980	2 G75523	probable cell divi
22	66	9.9	1171	2 T17454	diaphanous-related
23	65.5	9.8	501	2 T44939	glycerolaldehyde-3-p
24	65.5	9.8	562	2 S57951	beta-fructofuranos
25	65.5	9.8	710	2 T46589	ropy-2 protein [im
26	65	9.8	422	2 I37891	interleukin-11 rec
27	65	9.8	442	2 A57374	paired box transcr
28	65	9.8	464	1 JE0092	NADH dehydrogenase
29	65	9.8	566	1 T43706	isobutyryl-CoA mut

30 64.5 9.7 131 2 B75433  
31 64.5 9.7 339 2 S39979  
32 64.5 9.7 340 2 S40414  
33 64.5 9.7 396 2 S68772  
34 64.5 9.7 396 2 S68773  
35 64.5 9.7 437 2 S15704  
36 64 9.6 169 2 F70799  
37 64 9.6 342 2 S57612  
38 64 9.6 464 1 A39362  
39 64 9.6 472 2 T33800  
40 64 9.6 566 1 T46549  
41 64 9.6 622 2 H64447  
42 64 9.6 690 2 T31487  
43 64 9.6 705 1 C1HURB  
44 64 9.6 990 2 T14756  
45 64 9.6 1182 2 I48378

## ALIGNMENTS

## RESULT 1

E37831  
phenol 2-monooxygenase (EC 1.14.13.7) chain P4 - Pseudomonas sp. (strain CF600)  
C:Species: Pseudomonas sp.  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: E37831  
R:Nordlund, I.; Powlowski, J.; Shingler, V.  
J. Bacteriol. 172, 6826-6833, 1990  
A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol 2-monooxygenase component O  
A:Reference number: A37831; MUID:91072230  
A:Accession: E37831  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-119 <NOR>  
A:Cross-references: GB:M60276; GB:M37764; NID:g151449; PIDN:AAA25943.1; PID:g151454  
C:Superfamily: phenol 2-monooxygenase component O  
C:Keywords: oxidoreductase

Query Match 32.4%; Score 215.5; DB 1; Length 119;  
Best Local Similarity 37.6%; Pred. No. 1.1e-15;  
Matches 44; Conservative 18; Mismatches 54; Indels 1; Gaps 1;

QY 1 MAVIALKPYDFPVKDAVEKPPA-PLLYVCWENHLMPPAPCLPLPDMPFGALAGDVLPP 59  
Db 1 MTVNSIGETATPRDYOANFNGMQLLYLWEEHLMYCSALAFVLVAPGMPFAEFLEQLKP 60  
QY 60 VGYGHPDFAKIDWRVWFPSGEPWAPDPAKSLAGNGLGHKDLISFRTPLDGLGGA 116  
Db 61 ATHAHDSAKIDFSQALWQLNDQPTFPDYAASLEANGIDHKSMLRLNTPCLNGIQGS 117

## RESULT 2

S44307  
phenol 2-monooxygenase (EC 1.14.13.7) component O - Pseudomonas putida  
N:Alternate names: phenolhydroxylase chain E  
C:Species: Pseudomonas putida  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: F58972; S54764; S44307; S47418  
R:Ng, L.C.; Shingler, V.; Size, C.C.; Poh, C.L.  
Gene 151, 29-36, 1994

A:Title: Cloning and sequences of the first eight genes of the chromosomally encoded phenol 2-monooxygenase  
A:Reference number: A58972; MUID:95129877  
A:Accession: F58972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-119 <NGL>

A:Cross-references: EMBL:X79063; NID:g483477; PIDN:CAA55664.1; PID:g483482  
A:Experimental source: strain P35X (NCBI 9869)  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1994  
R:Herrmann, H.; Mueller, C.; Schmidt, I.; Mahne, J.; Petruschka, L.; Hahnke, K.  
Mol. Gen. Genet. 247, 240-246, 1995

A:Title: Complete nucleotide sequence of tbud, the gene encoding phenol/cresol hydroxylase  
A:Reference number: A45730; MUID:93015703  
A:Accession: A45730  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-671 <KUK>  
A:Cross-references: GB:M98806; NID:g151573; PIDN:AAA25992.1; PID:g151574  
C:Genetics:  
A:Gene: tbud  
C:Keywords: flavoprotein; oxidoreductase

Query Match	19.1%	Score 127;	DB 2;	Length 671;
Best Local Similarity	27.7%	Pred. NO. 1.5e-05;		
Matches 41;	Conservative 13;	Mismatches 28;	Indels 66;	Gaps 8;
Qy	1	MAVIA-LK-PYDPVKDAVEKEFAPL-----LYVCWENHLMFPA	37	
Db	202	MSVYTNLKAAPYKFLMKVDENFGELSTSGKKTTSIADESAIDAAJLHAWDAD	256	
Qy	38	PFCLPLPDMFPGALAGDVLPPVGYHPDFAK----IDWDRVEVFRSGEPWAPQPAKS--	91	
Db	257	-----DLGA-----AWHLDEASGLRAVDNNAQNFKSGQPTTPDAAKSLQ	296	
Qy	92	-----LAGNGLGHKDLISFRTPGLDGLG	114	
Db	297	EGRVFLAGDA-----RHRHPPLTGLG	317	

RESULT 5  
S46295  
Chlorophyll a/b-binding protein type II - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 01-Feb-1995 #sequence\_revision 19-Jan-1996 #text\_change 26-Aug-1999  
C/Accession: S46295  
R/Zhang, H.; Wang, J.; Goodman, H.M.  
Plant Mol. Biol. 25, 551-557, 1994  
A/Title: Differential expression in Arabidopsis of Lhca2, a PSI cab gene.  
A/Reference number: S46295; MUID:94325479  
A/Accession: S46295  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-271 <2HA>  
A/Cross-references: EMBL:U003395; NID:g541564; PIDN:AAA57542.1; PID:g541565  
C/Genetics:  
A/Gene: Lhca2; Cab5  
C/Superfamily: chlorophyll a/b-binding protein  
C/Keywords: photosystem I; transmembrane protein

```

Query Match      12.7%; Score 84.5; DB 2; Length 271;
Best Local Similarity 35.1%; Pred. No. 0.16;
Matches 20; Conservative 6; Mismatches 20; Indels 11; Gaps 2
QY 40 CLPPPPMPGALAG-----DWLPVYGYHPDFAKIDWRVFRSGE----PWA 85
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 CEPUPPPRGYSVALHNLWDLSGLPGDGFDPGLGSDPDPTLKKFAOALHHSWA 115
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT: 6  
S52261  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) flavoprotein 1 precursor - potato  
N;Alternate names: mitochondrial complex I 51K chain; NADH dehydrogenase (ubiquinone)  
C;Species: Solanum tuberosum (potato)  
C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 11-Jun-1999  
C;Accession: S52261  
R;Grohmann, L.; Thieck, O.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: S52261  
A;Accession: S52261  
A;Molecule type: mRNA  
A;Residues: 1-487 <GRO>  
A;Cross-references: EMBL:X83999; NID:g639833; PIDN:CAA58823.1; PID:g639834

A;Reference number: S45927

D47301

VlrB9 homolog - Bordetella pertussis  
C:Species: Bordetella pertussis  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
C:Accession: D47301  
R:Weiss, A.A.; Johnson, F.D.; Burns, D.L.  
Proc. Natl. Acad. Sci. U.S.A. 90, 2970-2974, 1993  
A:Title: Molecular characterization of an operon required for pertussis toxin secretion.  
A:Reference number: A47301; MUID:93219406  
A:Contents: BP338  
A:Accession: D47301  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-273 <WEI>  
A:Note: sequence extracted from NCBI backbone (NCBI:128775, NCBI:128780)

Query Match 10.7%; Score 71.5; DB 2; Length 273;  
Best Local Similarity 21.6%; Pred. No. 3.6;  
Matches 38; Conservative 20; Mismatches 47; Indels 71; Gaps 8;  
QY 3 VIALKPYDFPVKDAVEKFPAPLLYVCWENHLM----- 34  
DB 83 VLIKAKSFAQGSQAQPEGL---WNTLLVKTDRRLYDFDLVLASADAATPQALQRS 139  
QY 35 -----FP-----APFCLPLPPDPFGAL-----AGDVLPPVYGYHDPFAKI 70  
DB 140 RMAYRLQFVRYPAQAASRASPVGPAVAGALNRRYAMQVNGSDGIAPTAAY----- 192  
QY 71 DWRVVEW--FRSGEPW-----APDPAKSLAGNGLGHKDLISFR--TPGLDGLGGAS 117  
DB 193 DDCGRHTLWTFRPGQFPFAVFAVADPTGTETLVNLHIDNQSLVIRHAPVLMRLSGAS 248

RESULT 9  
S69546  
phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) precursor, mitochondrial - human  
C:Species: Homo sapiens (man)  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 22-Jun-1999  
C:Accession: S69546  
R:Wodareski, S.; Christ, B.; Bratke, J.; Zahn, S.; Heise, T.; Jungermann, K.  
Biochem. J. 315, 807-814, 1996  
A:Title: Molecular cloning, sequencing and expression of the cDNA of the mitochondrial  
A:Reference number: S69546; MUID:96220691  
A:Accession: S69546  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-640 <MOD>  
A:Cross-references: EMBL:X92720; NID:g1403049; PIDN:CAA63380.1; PID:e209003; PID:g140305  
A:Note: the authors translated the codon GCC for residue 477 as Arg  
C:Genetics:  
A:Genome: nuclear  
C:Superfamily: phosphoenolpyruvate carboxykinase (GTP)  
C:Keywords: carbon-carbon lyase; carboxy-lyase; mitochondrion

Query Match 10.7%; Score 71.5; DB 2; Length 640;  
Best Local Similarity 25.4%; Pred. No. 9.7;  
Matches 34; Conservative 16; Mismatches 51; Indels 33; Gaps 7;  
QY 6 LKPYDFPVKDAVEKFPAPLLYVCWENHLMFPAPFCLPLPPDPFGALAGDVLPPVY---- 61  
DB 135 MSPADPQ-RAVDSRFPG---CMQGRITMYLVLPFSMG-PVGSPLSRIGVQLTDSAYVVAS 187  
QY 62 -----GYHPDPAKIDWRVWFRS-----GEP---WAPDPAKSLAGNGLGHKDLI 103  
DB 188 MRITWRLTGPVLOALGDGFVKCLHSVGQPLTGQGPVSWQPCNPBKTLIHGVDPQREII 247  
QY 104 SFRTPGLDGLGGAS 117  
DB 248 SFGS-----GVGGNS 257  
RESULT 10

S52216  
viral proteinase - rabies virus  
C:Species: rabies virus  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Aug-1998  
C:Accession: S52216  
R:Camacho, A.; Tabaro, E.  
submitted to the EMBL Data Library, June 1994  
A:Reference number: S52215  
A:Accession: S52216  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-522 <CAM>  
A:Cross-references: EMBL:X79983  
A:Note: the source is designated as pseudorabies virus  
C:Superfamily: varicella-zoster virus gene 33 protein

Query Match 10.7%; Score 71; DB 2; Length 522;  
Best Local Similarity 23.4%; Pred. No. 8.7;  
Matches 22; Conservative 10; Mismatches 22; Indels 40; Gaps 3;  
QY 32 HLMFPAPFCLPLPPD-MPFGAL-----AGDVLPPVYGYHDPFAKIDWRVWFR 79  
DB 349 HVRYPTPYAPAPPQQLPPGAVGVGPHPHPHHAAGALYPPMYAPQGL----- 396  
QY 80 SGEPWAPDPAKSLAGNGLGHKDLISFRTPGLDGL 113  
DB 397 -----HAPPFSP-----VAHAVPALPGL 414

RESULT 11  
S26045  
sex-determining protein transformer - fruit fly (Drosophila hydei)  
C:Species: Drosophila hydei  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 17-Mar-1999  
C:Accession: S26045  
R:O'Neill, M.T.; Belote, J.M.  
Genetics 131, 113-128, 1992  
A:Title: Interspecific comparison of the transformer gene of drosophila reveals an un  
A:Reference number: S26044; MUID:92275311  
A:Accession: S26045  
A:Molecule type: DNA  
A:Residues: 1-201 <ONE>  
A:Cross-references: EMBL:X66931; NID:g7457; PID:g7458  
A:Note: the authors did not translate the codon for residue 111  
C:Genetics:  
A:Gene: tra  
A:Cross-references: FlyBase:FBgn0012404  
A:Introns: 11/3; 134/3

Query Match 10.2%; Score 68; DB 2; Length 201;  
Best Local Similarity 33.3%; Pred. No. 5.9;  
Matches 20; Conservative 7; Mismatches 27; Indels 6; Gaps 3;  
QY 35 FPAPFCLPLPPDMP-FGALAGDV---LP--PVYGYHDPFAKIDWRVWFRSGEPWAPDP 88  
DB 130 YPAYAWPPPPQAPQPNMYGAVPYGMPSPRYPAHYFAPYPRPLTTPYRAPPRPHP 189

RESULT 12  
T40809  
conserved hypothetical zinc-finger protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40809  
R:Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1998  
A:Reference number: Z21949  
A:Accession: T40809  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-482 <BEC>







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:09 ; Search time 58.11 Seconds  
(without alignments)  
62.927 Million cell updates/sec

Title: US-09-430-029-6  
Perfect score: 666  
Sequence: 1 MAVTALKPYDFPKDAVEKF.....HKDLISFRTPLDGLGGSF 118

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	215.5	32.4	119	1	DMPQ_PESP
2	127	19.1	670	1	TBUD_BURPI
3	74	11.1	1189	1	HAIR_HUMAN
4	72.5	10.9	495	1	TPS1_YEAST
5	71.5	10.7	286	1	FDRD_ALCEU
6	71.5	10.7	640	1	PPCM_HUMAN
7	68	10.2	201	1	TRSF_DROHY
8	67.5	10.1	405	1	LIM1_BRARE
9	67	10.1	1181	1	HAIR_RAT
10	66.5	10.0	345	1	CEBB_HUMAN
11	66.5	10.0	642	1	PHSA_STRAT
12	65	9.8	464	1	NUBM_HUMAN
13	65	9.8	491	1	PRCC_HUMAN
14	64.5	9.7	396	1	DPS1_PINST
15	64.5	9.7	396	1	DPS2_PINST
16	64.5	9.7	437	1	MYCN_XENLA
17	64.5	9.7	502	1	VL2_HPV48
18	64	9.6	342	1	P2_ARATH
19	64	9.6	443	1	HXA3_MOUSE
20	64	9.6	464	1	NUBM_BOVIN
21	64	9.6	705	1	C1R_HUMAN
22	64	9.6	1182	1	C1R_MOUSE
23	63.5	9.5	245	1	Y882_HAEIN
24	63.5	9.5	511	1	CP4B_HUMAN
25	63	9.5	515	1	ENG_BLV
26	63	9.5	529	1	GAG_MLVDU
27	63	9.5	1048	1	SRA4_RAT
28	62.5	9.4	345	1	P1_ARATH
29	62.5	9.4	382	1	LMX1_MESAU
30	62.5	9.4	736	1	DVL2_HUMAN
31	62	9.3	420	1	PHR_THETH
32	61.5	9.2	397	1	ASSY_METVA
33	61.5	9.2	493	1	YLD2_CAEL

34 61.5 9.2 920 1 PARC\_SVNY3 P73077 synchocyst  
35 61.5 9.2 1071 1 UB7\_YEAST P40453 saccharomyc  
36 61.5 9.2 1458 1 PHLX\_RABIT Q05017 oryctolagus  
37 61 9.2 257 1 NTGA\_HUMAN P34132 homo sapien  
38 61 9.2 540 1 YFE0\_YEAST P43562 saccharomyc  
39 61 9.2 541 1 YR52\_HUMAN O00482 homo sapien  
40 61 9.2 560 1 CALX\_SCHPO P36581 schizosacch  
41 60.5 9.1 268 1 CDX1\_MOUSE P18111 mus musculus  
42 60.5 9.1 277 1 HMPH\_CHICK Q05502 gallus gall  
43 60.5 9.1 376 1 PROB\_THETH Q60050 thermus aqu  
44 60.5 9.1 517 1 DHA5\_HUMAN P30837 homo sapien  
45 60 9.0 405 1 YNX4\_CAEL P34577 caenorhabdi

#### ALIGNMENTS

RESULT 1  
DMPQ\_PESP  
ID DMPQ\_PESP STANDARD; PRT; 119 AA.  
AC P19733;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PHENOL HYDROXYLASE P4 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE  
DE P4 COMPONENT)  
GN DMPQ OR PHEA5.  
OS Pseudomonas sp. (strain CF600).  
OG Plasmid pVil150.  
OC Bacteria; Proteobacteria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91072230.  
RA Nordlund I., Powlowski J., Shingler V.;  
RT "Complete nucleotide sequence and polypeptide analysis of  
RT multicomponent phenol hydroxylase from Pseudomonas sp. strain  
RT CF600".  
RL J. Bacteriol. 172:6826-6833(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BH;  
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,  
RA Fujita M.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED  
CC DERIVATIVES. P4 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR  
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.  
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) -> CATECHOL + NADP(+) +  
CC H(2)O.  
CC -!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.  
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.  
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED  
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M60276; AAA25943.1; -.  
DR EMBL; D28864; BAA06018.1; -.  
DR PIR; E37831; E37831.  
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;  
KW NADP; Flavoprotein; FAD; Iron; Plasmid.  
SQ SEQUENCE 119 AA; 13207 MW; AE0151A918638C49 CRC64;

Query Match 32.4%; Score 215.5; DB 1; Length 119;  
Best Local Similarity 37.6%; Pred. No. 2.5e-15;  
Matches 44; Conservative 18; Mismatches 54; Indels 1; Gaps 1;

RESULT	3	HAIR_HUMAN	STANDARD;	PRT: 1189 AA.
ID	AC	043593;		
DT	DT	15-JUL-1999 (Rel. 38, Created)		
DT	DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	DE	HAIRLESS PROTEIN.		
HS	HS			
OS	OS	Homo sapiens (Human).		
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
RP	RP	[1]		
RP	RP	SEQUENCE FROM N.A., AND VARIANT ALUNC ALA-1022.		
RX	RX	MEDLINE; 9811413.		
RA	RA	Ahmad W., Ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,		
RA	RA	Lam H., Alta V.M., Owen J., Deblaquiere M., Frank J.,		
RA	RA	Cserhalmi-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,		
RA	RA	Ahmad M., Ott J., Christiano A.M.;		
RT	RT	"Alopecia universalis associated with a mutation in the human hairless		
RT	RT	gene.;" 279:720-724(1998).		
RL	RL	[2]		
RP	RP	VARIANT APL GLN-620.		
RP	RP	MEDLINE; 98431781.		
RX	RX	Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,		
RA	RA	Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;		
RA	RA	"A missense mutation in the zinc-finger domain of the human hairless		
RT	RT	gene underlies congenital atrichia in a family of Irish travellers.;"		
RT	RT	Am. J. Hum. Genet. 63:984-991(1998).		
RL	RL	[3]		
RP	RP	VARIANT ALUNC ASP-1136, AND ALTERNATIVE SPLICING.		
RP	RP	MEDLINE; 98409496.		
RX	RX	Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,		
RA	RA	Farooq S.A., Al-Hafiri K.S., Ahmad M., Haque S., Rietschel M.,		
RA	RA	Propping P., Kruse R., Nothen M.M.;		
RT	RT	"Cloning, genomic organization, alternative transcripts and mutational		
RT	RT	analysis of the gene responsible for autosomal recessive universal		
RT	RT	congenital alopecia.;"		
RL	RL	Hum. Mol. Genet. 7:1671-1679(1998).		
CC	CC	-!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO		
CC	CC	REGULATE ONE OF THE PHASES OF HAIR GROWTH.		
CC	CC	-!- SUBCELLULAR LOCATION: NUCLEAR.		
CC	CC	-!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY		
CC	CC	ALTERNATIVE SPLICING.		
CC	CC	-!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND SKIN.		
CC	CC	-!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS		
CC	CC	(ALUNC); A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS		
CC	CC	CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.		
CC	CC	-!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHIA WITH PAPULAR		
CC	CC	LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHIA) THIS AUTOSOMAL		
CC	CC	RECESSIVE DISEASE IS CHARACTERIZED BY PAPILLARY LESIONS OVER MOST		
CC	CC	OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.		
CC	CC	-----		
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collabora		
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	CC	the European Bioinformatics Institute. There are no restrictions on		
CC	CC	use by non-profit institutions as long as its content is in no		
CC	CC	modified and this statement is not removed. Usage by and for commerc		
CC	CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announc">http://www.isb-sib.ch/announc</a>		
CC	CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	CC	-----		
DR	DR	EMBL; AF039196; AAC32258.1; -		
DR	DR	MIM; 602302; -		
DR	DR	MIM; 203655; -		
DR	DR	MIM; 209500; -		
KW	KW	zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;		
KW	KW	Metal-binding; Alternative splicing; Disease mutation.		
FT	FT	ZN_FING 600 625		
FT	FT	VARIANT 620 620		
FT	FT	R -> Q (IN APL).		
FT	FT	/FTId=VAR_005265.		